

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number  
WO 01/46397 A2

(51) International Patent Classification<sup>7</sup>: C12N 9/12

(21) International Application Number: PCT/US00/35304

(22) International Filing Date:  
20 December 2000 (20.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/172,066 23 December 1999 (23.12.1999) US  
60/176,107 14 January 2000 (14.01.2000) US  
60/177,731 21 January 2000 (21.01.2000) US  
60/178,573 28 January 2000 (28.01.2000) US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): YANG, Junming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). REDDY, Roopa [IN/US]; 1233 W. McKinley Avenue #3,

Sunnyvale, CA 94086 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). KHAN, Farrah, A. [IN/US]; 333 Escuela Avenue #221, Mountain View, CA 94040 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/46397 A2

(54) Title: HUMAN KINASES

(57) Abstract: The invention provides human kinases (PKIN) and polynucleotides which identify and encode PKIN. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PKIN.

**HUMAN KINASES****TECHNICAL FIELD**

This invention relates to nucleic acid and amino acid sequences of human kinases and to the use  
5 of these sequences in the diagnosis, treatment, and prevention of cancer, immune disorders, disorders  
affecting growth and development, cardiovascular diseases, and lipid disorders, and in the assessment of  
the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of  
human kinases.

10

**BACKGROUND OF THE INVENTION**

Kinases comprise the largest known enzyme superfamily and vary widely in their target  
molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to a  
phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most  
kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of  
15 molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are  
phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the  
acceptor molecule, causing internal conformational changes and potentially influencing intermolecular  
contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in  
eukaryotic cells. In general, proteins are activated by phosphorylation in response to extracellular  
20 signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated  
proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second  
messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens,  
that regulate protein phosphorylation.

Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as  
25 glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular  
environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells  
has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle  
have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked to  
diseases and disorders of the reproductive system, immune system, and skeletal muscle.

30 There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs),  
phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs),  
phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics  
of both families and have dual specificity for both tyrosine and serine/threonine residues. Almost all  
kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and

sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and Hanks, S. (1995) The Protein Kinase Facts Book, Vol I p.p. 17-20 Academic Press, San Diego, CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

## 20 Protein Tyrosine Kinases

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased

tyrosine phosphorylation activity (Charbonneau, H. and Tonks, N. K. (1992) *Annu. Rev. Cell Biol.* 8:463-93). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Protein Serine/Threonine Kinases

5       Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent 10 the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include 15 the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

      The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 20 3,4,5-triphosphate, cyclic ADP-ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, 25 bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York, NY, pp. 416-431, 1887).

30       The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism, and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al., (1995) *J. Biol. Chem.* 270:14875-14883.

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al., (1998) *J. Biol. Chem.* 273:1357-1364.) The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$ . Fish et al., identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al, *supra*.)

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state. Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al., have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al., (2000) *Science* 288:483-491.)

#### Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) *EMBO Journal* 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through

phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) *BioEssays* 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be enriched in mammalian forebrain. This protein is associated with vesicles in both axons and dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to 5 CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) *J. Neurosci.* 14:1-13).

#### Mitogen-Activated Protein Kinases

The mitogen-activated protein kinases (MAP) which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades are another STK family that regulates intracellular 10 signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and Weinberg, R.A. (1993) *Nature* 365:781-783). MAP kinase signaling pathways are present in mammalian cells as well as in yeast. The extracellular stimuli which activate MAP kinase pathways include epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, endotoxic lipopolysaccharide (LPS), and 15 pro-inflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.

#### Cyclin-Dependent Protein Kinases

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells 20 through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the 25 dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M., et al. (1998) *EMBO J.* 17:470-481).

#### Checkpoint and Cell Cycle Kinases

In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is

induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the 5 G2/M transition. (Sanchez, Y. et al. (1997) *Science* 277:1497-1501.) Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis. (Peng, C-Y et al. (1997) *Science* 277:1501- 1505.) Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A similar deficiency in a checkpoint kinase, such as 10 Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

#### Proliferation-Related Kinases

15 Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) *J. Biol. Chem.* 271:19402-8). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

#### 5'-AMP-activated protein kinase

20 A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) *J. Biol. Chem.* 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of 25 a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

#### Kinases in Apoptosis

30 Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune disease, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This C-terminal domain appears to

mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) *J. Biol. Chem.*, 273:29066-29071). DRAK1 and DRAK2 are STKs that share homology with the death-associated protein kinases (DAP kinases), known to function in interferon- $\gamma$  induced apoptosis (Sanjo et al. *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al. *supra*). However, deletion of either the N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) *J. Biol. Chem.* 273:12296-12300). CD95 is a member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) *Cell* 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al. *supra*).

#### Mitochondrial Protein Kinases

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) *Adv. Enzyme Regul.* 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to a branched-

chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) *Adv. Enzyme Regul.* 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) *supra*).

### KINASES WITH NON-PROTEIN SUBSTRATES

10 Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) *Curr. Opin. Cell. Biol.* 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP<sub>2</sub>). PIP<sub>2</sub> is then cleaved into inositol triphosphate (IP<sub>3</sub>) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. 25 PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) bisphosphate (PIP<sub>2</sub>) to PI (3,4,5) P<sub>3</sub> (PIP<sub>3</sub>). PIP<sub>3</sub> then activates a number of other proteins, including 30 PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R., et al. (1998) *Biochem. J.* 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in

type-2 diabetes (Shepard, *supra*).

An example of lipid kinase phosphorylation activity is the phosphorylation of D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, 5 T. et al. (1998) *J. Biol. Chem.* 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including 10 platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al. *supra*).

#### Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and 15 guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs 20 for cancer and antiviral therapy (Pillwein, K. et al. (1990) *Cancer Res.* 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring 25 high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zeleznikar, R.J. et al. (1995) *J. Biol. Chem.* 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy 30 disorders that can result in cardiac or respiratory failure and may be treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell

receptors, subsequent intracellular activation of adenyl cyclase, and production of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and GTP levels also control the activity of certain oncogenic proteins such as p21<sup>ras</sup> known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) *Cancer Res.* 49:4682-4689). High ratios of 5 GTP:GDP caused by suppression of GuK cause activation of p21<sup>ras</sup> and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir 10 and buclovir (Miller, W.H. and Miller R.L. (1980) *J. Biol. Chem.* 255:7204-7207; Stenberg, K. et al. (1986) *J. Biol. Chem.* 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

#### Pyrimidine Kinases

15 The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several 20 chemotherapeutically important nucleoside analogues (Arner E.S. and Eriksson, S. (1995) *Pharmacol. Ther.* 67:155-186).

The discovery of new human kinases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and 25 lipid disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human kinases.

#### **SUMMARY OF THE INVENTION**

The invention features purified polypeptides, human kinases, referred to collectively as "PKIN" 30 and individually as "PKIN-1," "PKIN-2," "PKIN-3," "PKIN-4," "PKIN-5," "PKIN-6," "PKIN-7," "PKIN-8," "PKIN-9," "PKIN-10," "PKIN-11," and "PKIN-12." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the

group consisting of SEQ ID NO:1-12, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-12.

5 The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of an amino acid sequence selected from the group consisting of 10 SEQ ID NO:1-12, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-12. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:13-24.

15 Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID 20 NO:1-12, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

25 The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of an amino acid sequence selected from the group 30 consisting of SEQ ID NO:1-12. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a

polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of an amino acid sequence selected 5 from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a naturally occurring polynucleotide sequence having at least 90% sequence 10 identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, 15 said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) 20 an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and 25 optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of 30 SEQ ID NO:13-24, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or

absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising

an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

5 The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of an amino acid sequence selected  
10 from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

15 The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of an amino acid sequence  
20 selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with  
25 the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a  
30 sequence selected from the group consisting of SEQ ID NO:13-24, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

20

#### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for each polypeptide of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of each polypeptide sequence, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of each polypeptide.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble each polynucleotide sequence, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for each polynucleotide of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

## DESCRIPTION OF THE INVENTION

5 Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

10 It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

15 Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, 20 protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

25 "PKIN" refers to the amino acid sequences of substantially purified PKIN obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PKIN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with 30 PKIN or by acting on components of the biological pathway in which PKIN participates.

An "allelic variant" is an alternative form of the gene encoding PKIN. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to

allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides.

Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding PKIN include those sequences with deletions,

5 insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PKIN or a polypeptide with at least one functional characteristic of PKIN. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PKIN, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PKIN. The 10 encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PKIN. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PKIN is retained. For example, negatively charged amino acids may 15 include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

20 The terms “amino acid” and “amino acid sequence” refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where “amino acid sequence” is recited to refer to a sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

25 “Amplification” relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of PKIN. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small 30 molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant.

Antibodies that bind PKIN polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers 5 that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which 10 bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; 15 peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once 20 introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical 25 functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PKIN, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid 30 sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PKIN or fragments of PKIN may be

employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

5 "Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap 10 (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

15 "Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as 20 conservative amino acid substitutions.

	Original Residue	Conservative Substitution
20	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
25	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
30	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
35	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, 40 (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, 5 hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a 10 measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of PKIN or the polynucleotide encoding PKIN which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment 15 used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain 20 defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:13-24 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:13-24, for example, as distinct from any other sequence in the 25 genome from which the fragment was obtained. A fragment of SEQ ID NO:13-24 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:13-24 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:13-24 and the region of SEQ ID NO:13-24 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

30 A fragment of SEQ ID NO:1-12 is encoded by a fragment of SEQ ID NO:13-24. A fragment of SEQ ID NO:1-12 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-12. For example, a fragment of SEQ ID NO:1-12 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-12. The precise length of a fragment of SEQ ID NO:1-12 and the region of SEQ ID NO:1-12 to which the fragment 35 corresponds are routinely determinable by one of ordinary skill in the art based on the intended

purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

5 "Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a 10 standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore 15 achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default 20 parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between 25 aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at 25 <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The 30 "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

5       *Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over 10 the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

15       Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

20       The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

25       Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default residue weight table. As with 30 polynucleotide alignments, the percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

5 *Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, 10 a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

15 “Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

20 “Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency 25 of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive 30 annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic

strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking 10 reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200  $\mu$ g/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency 15 conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_{0t}$  or  $R_{0t}$  analysis) or formed between one nucleic acid 20 sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

25 "Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PKIN which is 30 capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PKIN which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PKIN. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, 5 functional, or immunological properties of PKIN.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

10 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

15 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

20 "Post-translational modification" of an PKIN may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PKIN.

25 "Probe" refers to nucleic acid sequences encoding PKIN, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA 30 polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100,

or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence

that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have 5 been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is 10 expressed, inducing a protective immunological response in the mammal.

A “regulatory element” refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

15 “Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An “RNA equivalent,” in reference to a DNA sequence, is composed of the same linear 20 sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term “sample” is used in its broadest sense. A sample suspected of containing PKIN, nucleic acids encoding PKIN, or fragments thereof may comprise a bodily fluid; an extract from a cell, 25 chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms “specific binding” and “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure 30 of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term “substantially purified” refers to nucleic acid or amino acid sequences that are

removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by 5 different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

10 A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences 15 into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed 20 cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor 25 of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be 30 introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at

least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule.

5 Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be

10 indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

15

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human kinases (PKIN), the polynucleotides 25 encoding PKIN, and the use of these compositions for the diagnosis, treatment, or prevention of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide 30 sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 5 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of each of the polypeptides of the invention. Columns 10 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). 15 Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any 20 combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification 25 technologies that identify SEQ ID NO:13-24 or that distinguish between SEQ ID NO:13-24 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') 30 and stop (3') positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 2287966H1 is the identification number of an Incyte cDNA sequence, and BRAINON01 is the cDNA library from which

it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70166939V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g2821547) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, g4454511.v113.gs\_3.nt.edit is the identification number of a Genscan-predicted coding sequence, with g4454511 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon stitching” algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon-stretching” algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

15 Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

20 The invention also encompasses PKIN variants. A preferred PKIN variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PKIN amino acid sequence, and which contains at least one functional or structural characteristic of PKIN.

25 The invention also encompasses polynucleotides which encode PKIN. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:13-24, which encodes PKIN. The polynucleotide sequences of SEQ ID NO:13-24, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

30 The invention also encompasses a variant of a polynucleotide sequence encoding PKIN. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PKIN. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:13-24 which has at

least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:13-24. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PKIN.

5 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PKIN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in  
10 accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PKIN, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PKIN and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PKIN under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PKIN or its  
15 derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PKIN and its derivatives without altering the encoded amino acid sequences include the production of RNA  
20 transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PKIN and PKIN derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using  
25 reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PKIN or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:13-24 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and  
30 S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of

DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is 5 automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a 10 variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PKIN may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, 15 such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a 20 known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and 25 ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in 30 finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been

size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

5 Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate 10 software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

15 In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PKIN may be cloned in recombinant DNA molecules that direct expression of PKIN, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PKIN.

20 The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PKIN-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, 25 alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. et al. (1999) *Nat. Biotechnol.* 17:259-264; and Crameri, A. et al. (1996) *Nat. Biotechnol.* 14:315-319) to alter or 30 improve the biological properties of PKIN, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of

DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of 5 homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PKIN may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids 10 Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PKIN itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp.55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be 15 achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PKIN, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid 20 chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PKIN, the nucleotide sequences encoding PKIN or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains 25 the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PKIN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PKIN. Such signals 30 include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PKIN and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the

vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

5 Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PKIN and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) 10 *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PKIN. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); 15 plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di 20 Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) 25 The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending 30 upon the use intended for polynucleotide sequences encoding PKIN. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PKIN can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PKIN into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed

bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of PKIN are needed, e.g. for the production of antibodies, 5 vectors which direct high level expression of PKIN may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PKIN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such 10 vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of PKIN. Transcription of sequences encoding 15 PKIN may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can 20 be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases 25 where an adenovirus is used as an expression vector, sequences encoding PKIN may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PKIN in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV- 30 based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PKIN in cell lines is preferred. For example, sequences encoding PKIN can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the 5 introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

10 Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers 15 resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 20:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  25 glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

30 Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PKIN is inserted within a marker gene sequence, transformed cells containing sequences encoding PKIN can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PKIN under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PKIN and that express PKIN may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and

protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PKIN using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include 5 enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PKIN is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. 10 et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PKIN include oligolabeling, 15 nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PKIN, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available 20 kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PKIN may be cultured under 25 conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PKIN may be designed to contain signal sequences which direct secretion of PKIN through a prokaryotic or eukaryotic cell membrane.

30 In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells

which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

5 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PKIN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PKIN protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PKIN activity. Heterologous protein and peptide moieties 10 may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, 15 respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PKIN encoding sequence and the heterologous protein sequence, so that PKIN may be cleaved away from the heterologous moiety following purification. Methods for fusion protein 20 expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PKIN may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or 25 SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

PKIN of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PKIN. At least one and up to a plurality of test compounds may be screened for specific binding to PKIN. Examples of test compounds include antibodies, oligonucleotides, proteins 30 (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PKIN, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) *Current Protocols in Immunology* 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PKIN

binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PKIN, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PKIN or cell membrane fractions which contain PKIN are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PKIN or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, 10 the assay may comprise the steps of combining at least one test compound with PKIN, either in solution or affixed to a solid support, and detecting the binding of PKIN to the compound.

Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a 15 solid support.

PKIN of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PKIN. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PKIN activity, wherein PKIN is combined with at least one test compound, and the activity of PKIN in the 20 presence of a test compound is compared with the activity of PKIN in the absence of the test compound.

A change in the activity of PKIN in the presence of the test compound is indicative of a compound that modulates the activity of PKIN. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising PKIN under conditions suitable for PKIN activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PKIN may do so 25 indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PKIN or their mammalian homologs may be “knocked out” in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of 30 human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by

homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330).

Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

5 Polynucleotides encoding PKIN may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell 10 lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

15 Polynucleotides encoding PKIN can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PKIN is injected into animal ES cells, and the injected sequence integrates into 20 the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PKIN, e.g., by secreting PKIN in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

## **THERAPEUTICS**

25 Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PKIN and human kinases. In addition, the expression of PKIN is closely associated with cancers, cell proliferation and cardiovascular diseases. Therefore, PKIN appears to play a role in cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders. In the treatment of disorders associated with increased PKIN expression or activity, it is desirable to decrease the expression or activity of PKIN. In the treatment of disorders associated with decreased PKIN expression or activity, it is desirable to increase the expression or activity of PKIN.

30 Therefore, in one embodiment, PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal

tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, 5 anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, 10 hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, 15 fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of 20 the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and 25 mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, 30 hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular

calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary

5 congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative

10 interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung

15 transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus,

20 lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease,

25 hyperlipidemia, hyperlipemia, lipid myopathies, and obesity.

In another embodiment, a vector capable of expressing PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PKIN in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PKIN may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity

of PKIN including, but not limited to, those listed above.

In a further embodiment, an antagonist of PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN. Examples of such disorders include, but are not limited to, those cancers, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders described above. In one aspect, an antibody which specifically binds PKIN may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PKIN.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PKIN may be produced using methods which are generally known in the art. In particular, purified PKIN may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PKIN. Antibodies to PKIN may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PKIN or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PKIN have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at

least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PKIN amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

5        Monoclonal antibodies to PKIN may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and 10 Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, 15 S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PKIN-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

20        Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

25        Antibody fragments which contain specific binding sites for PKIN may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

30        Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PKIN and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two

non-interfering PKIN epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PKIN. Affinity is expressed as an association 5 constant,  $K_a$ , which is defined as the molar concentration of PKIN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PKIN epitopes, represents the average affinity, or avidity, of the antibodies for PKIN. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular PKIN epitope, 10 represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the PKIN-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately 15 require dissociation of PKIN, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg 20 specific antibody/ml, is generally employed in procedures requiring precipitation of PKIN-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PKIN, or any fragment or 25 complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PKIN. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be 30 designed from various locations along the coding or control regions of sequences encoding PKIN. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totowa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g.,

Slater, J.E. et al. (1998) *J. Allergy Cli. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other 5 gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PKIN may be used for 10 somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), 15 cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated 20 cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the 25 case where a genetic deficiency in PKIN expression or regulation causes disease, the expression of PKIN from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PKIN are treated by constructing mammalian expression vectors encoding PKIN and introducing these vectors 30 by mechanical means into PKIN-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of PKIN include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PKIN may be expressed 5 using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the 10 ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PKIN from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID 15 TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these 20 standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PKIN expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PKIN under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive 25 element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for 30 receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a

method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference.

Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PKIN to cells which have one or more genetic abnormalities with respect to the expression of PKIN. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PKIN to target cells which have one or more genetic abnormalities with respect to the expression of PKIN. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PKIN to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 69:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary

skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PKIN to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PKIN into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PKIN-coding RNAs and the synthesis of high levels of PKIN in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of PKIN into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PKIN.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by

scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of 5 candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis.

10 Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PKIN. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible 15 modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterate linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, 20 guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PKIN. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, 25 transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PKIN expression or activity, a compound which specifically inhibits expression of the polynucleotide 30 encoding PKIN may be therapeutically useful, and in the treatment of disorders associated with decreased PKIN expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PKIN may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method

commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a 5 library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PKIN is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PKIN are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected 10 by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PKIN. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the 15 polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial 20 library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken 25 from the patient and clonally propagated for autologous transplant back into that same patient.

Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such 30 therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various

formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PKIN, antibodies to PKIN, and mimetics, agonists, antagonists, or inhibitors of PKIN.

The compositions utilized in this invention may be administered by any number of routes 5 including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. 10 These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting 15 formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of 20 macromolecules comprising PKIN or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PKIN or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) 25 Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for 30 administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PKIN or fragments thereof, antibodies of PKIN, and agonists, antagonists or inhibitors of PKIN, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by

calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD<sub>50</sub>/ED<sub>50</sub> ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are 5 used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject 10 requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending 15 on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu$ g to 100,000  $\mu$ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their 20 inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind PKIN may be used for the diagnosis 25 of disorders characterized by expression of PKIN, or in assays to monitor patients being treated with PKIN or agonists, antagonists, or inhibitors of PKIN. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PKIN include methods which utilize the antibody and a label to detect PKIN in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter 30 molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PKIN, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PKIN expression. Normal or standard values for PKIN expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PKIN under

conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PKIN expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

5 In another embodiment of the invention, the polynucleotides encoding PKIN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PKIN may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of  
10 PKIN, and to monitor regulation of PKIN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PKIN or closely related molecules may be used to identify nucleic acid sequences which encode PKIN. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a  
15 conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PKIN, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PKIN encoding sequences. The hybridization probes of the subject  
20 invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:13-24 or from genomic sequences including promoters, enhancers, and introns of the PKIN gene.

Means for producing specific hybridization probes for DNAs encoding PKIN include the cloning of polynucleotide sequences encoding PKIN or PKIN derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to  
25 synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PKIN may be used for the diagnosis of disorders associated  
30 with expression of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple

myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy

5 (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis,

10 osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis,

15 arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver,

20 lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary

25 keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and

30 phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial

thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary

5 hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis

10 obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine

15 deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid

20 adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity. The polynucleotide sequences encoding

25 PKIN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PKIN expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PKIN may be useful in assays that 30 detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PKIN may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control

sample then the presence of altered levels of nucleotide sequences encoding PKIN in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

5 In order to provide a basis for the diagnosis of a disorder associated with expression of PKIN, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PKIN, under conditions suitable for hybridization or amplification.

Standard hybridization may be quantified by comparing the values obtained from normal subjects with 10 values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, 15 hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or 20 overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

25 Additional diagnostic uses for oligonucleotides designed from the sequences encoding PKIN may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PKIN, or a fragment of a polynucleotide complementary to the polynucleotide encoding PKIN, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may 30 also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease

in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PKIN include radiolabeling or biotinylation of nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PKIN, fragments of PKIN, or antibodies specific for PKIN may be

used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is

important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PKIN to quantify the levels of PKIN expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoza, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of

methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation 5 between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

10        In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound 15 in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

20        In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

25        Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

30        In another embodiment of the invention, nucleic acid sequences encoding PKIN may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal

mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357.)

10       Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PKIN on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA  
15       associated with that disorder and thus may further positional cloning efforts.

*In situ* hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is  
20       valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may  
25       also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

      In another embodiment of the invention, PKIN, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a  
30       solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PKIN and the agent being tested may be measured.

      Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are

synthesized on a solid substrate. The test compounds are reacted with PKIN, or fragments thereof, and washed. Bound PKIN is then detected by methods well known in the art. Purified PKIN can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

5 In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PKIN specifically compete with a test compound for binding PKIN. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PKIN.

10 In additional embodiments, the nucleotide sequences which encode PKIN may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

15 Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/172,066, U.S. Ser. No. 60/176,107, U.S. Ser. No. 60/177,731, and U.S. Ser. No. 60/178,573, are expressly incorporated by reference herein.

20

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. The Incyte cDNA shown for SEQ 25 ID NO:13 was derived from a cDNA library constructed from musculoskeletal tissue. The Incyte cDNA shown for SEQ ID NO:14 was derived from cDNA libraries constructed from prostate, brain and ovarian tissues, including tissues associated with brain, prostate and thyroid tumors. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic 30 solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated

using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

5 In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic  
10 oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g.,  
15 PBLUESCRIPT plasmid (Stratagene), PSSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

20 **II. Isolation of cDNA Clones**

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 25 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4 °C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) *Anal. Biochem.* 216:1-14). Host cell lysis and thermal 30 cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

**III. Sequencing and Analysis**

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the

5 MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI  
10 PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

15 The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS,  
20 DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank  
25 cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length  
30 polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using

MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent 5 identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of 10 which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and 15 polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:13-24. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### **IV. Identification and Editing of Coding Sequences from Genomic DNA**

Putative human kinases were initially identified by running the Genscan gene identification 20 program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a 25 FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode human kinases, the encoded polypeptides were analyzed by querying against PFAM models for kinases. Potential human kinases were also identified by homology to Incyte cDNA sequences that had been annotated as kinases. These selected Genscan-predicted sequences were then 30 compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information

was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

5

#### **V. Assembly of Genomic Sequence Data with cDNA Sequence Data**

##### **"Stitched" Sequences**

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped 10 to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence.

Sequence intervals in which the entire length of the interval was present on more than one sequence in 15 the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the 20 longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit 25 from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

##### **"Stretched" Sequences**

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public 30 databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the

chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to 5 determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of PKIN Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:13-24 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched 10 SEQ ID NO:13-24 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment 15 of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, or human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in 20 humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified 25 disease genes map within or in proximity to the intervals indicated above.

#### **VII. Analysis of Polynucleotide Expression**

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) 30 supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar.

The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

5

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by

10 assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter  
15 15 of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PKIN are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least  
20 in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory  
25 system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all  
30 categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PKIN. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

### **VIII. Extension of PKIN Encoding Polynucleotides**

Full length polynucleotide sequences were also produced by extension of an appropriate

fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 5 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68 °C to about 72 °C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

10 High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg<sup>2+</sup>, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer 15 pair PCI A and PCI B: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 60 °C, 1 min; Step 4: 68 °C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68 °C, 5 min; Step 7: storage at 4 °C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 57 °C, 1 min; Step 4: 68 °C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68 °C, 5 min; Step 7: storage at 4 °C.

20 The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the 25 concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For 30 shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing

media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 5 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC 10 DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

15 **IX. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:13-24 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National 20 Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of  $[\gamma-^{32}\text{P}]$  adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per 25 minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature 30 under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

**X. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing

photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure 5 analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 10 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the 15 biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on 20 the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is 25 reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with 30 GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated

using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation

5 Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia

10 Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and 15 coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

20 Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

25 Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### Hybridization

30 Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

35 **Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide 5 containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, 10 Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously. 15 The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different 20 fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC 25 computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

30 A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

**XI. Complementary Polynucleotides**

Sequences complementary to the PKIN-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PKIN. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with 5 smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PKIN. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PKIN-encoding transcript.

**10 XII. Expression of PKIN**

Expression and purification of PKIN is achieved using bacterial or virus-based expression systems. For expression of PKIN in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the 15 T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element.

Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic 20 resistant bacteria express PKIN upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PKIN in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PKIN by either homologous recombination or bacterial-mediated transposition involving transfer plasmid 25 intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional 30 genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PKIN is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton 30 enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PKIN at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-

His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PKIN obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII where applicable.

### 5 XIII. Functional Assays

PKIN function is assessed by expressing the sequences encoding PKIN at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which 10 contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish 15 transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of 20 fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated 25 Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PKIN on gene expression can be assessed using highly purified populations of 30 cells transfected with sequences encoding PKIN and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PKIN and other genes of interest can be analyzed by northern analysis or microarray techniques.

**XIV. Production of PKIN Specific Antibodies**

PKIN substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

5        Alternatively, the PKIN amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

10        Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide 15 and anti-PKIN activity by, for example, binding the peptide or PKIN to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

**XV. Purification of Naturally Occurring PKIN Using Specific Antibodies**

Naturally occurring or recombinant PKIN is substantially purified by immunoaffinity chromatography using antibodies specific for PKIN. An immunoaffinity column is constructed by 20 covalently coupling anti-PKIN antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PKIN are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PKIN (e.g., high ionic strength buffers in the 25 presence of detergent). The column is eluted under conditions that disrupt antibody/PKIN binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PKIN is collected.

**XVI. Identification of Molecules Which Interact with PKIN**

PKIN, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent. 30 (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PKIN, washed, and any wells with labeled PKIN complex are assayed. Data obtained using different concentrations of PKIN are used to calculate values for the number, affinity, and association of PKIN with the candidate molecules.

Alternatively, molecules interacting with PKIN are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PKIN may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) 5 which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

## XVII. Demonstration of PKIN Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein 10 substrate by PKIN in the presence of gamma-labeled  $^{32}\text{P}$ -ATP. PKIN is incubated with the protein substrate,  $^{32}\text{P}$ -ATP, and an appropriate kinase buffer. The  $^{32}\text{P}$  incorporated into the substrate is separated from free  $^{32}\text{P}$ -ATP by electrophoresis and the incorporated  $^{32}\text{P}$  is counted using a radioisotope counter. The amount of incorporated  $^{32}\text{P}$  is proportional to the activity of PKIN. A determination of 15 the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma  $^{32}\text{P}$ -ATP. Following the reaction, free avidin in solution is added for binding to the 20 biotinylated  $^{32}\text{P}$ -peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma  $^{32}\text{P}$ -ATP. The reservoir of the centrifuged unit containing the  $^{32}\text{P}$ -peptide product as retentate is then counted in a scintillation counter. This procedure allows assay of any type of protein kinase 25 sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes are as follows: Histone H1 (Sigma) and p34<sup>cdc2</sup>kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) *Methods in Enzymology* 200:62-81).

30 In another alternative, protein kinase activity of PKIN is demonstrated *in vitro* in an assay containing PKIN, 50 $\mu\text{l}$  of kinase buffer, 1 $\mu\text{g}$  substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10  $\mu\text{g}$  ATP, and 0.5 $\mu\text{Ci}$  [ $\gamma$ - $^{33}\text{P}$ ]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [ $\gamma$ - $^{33}\text{P}$ ]ATP is removed by washing and the incorporated radioactivity is measured using a radioactivity scintillation

counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and visualized on a 12% SDS polyacrylamide gel by autoradiography. Incorporated radioactivity is corrected for reactions carried out in the absence of PKIN or in the presence of the inactive kinase, K38A.

5 In yet another alternative, adenylate kinase or guanylate kinase activity may be measured by the incorporation of  $^{32}\text{P}$  from gamma-labeled  $^{32}\text{P}$ -ATP into ADP or GDP using a gamma radioisotope counter. The enzyme, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and  $^{32}\text{P}$ -labeled ATP as the phosphate donor. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized  
10 and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is cut out and counted. The radioactivity recovered is proportional to the enzyme activity.

15 In yet another alternative, other assays for PKIN include scintillation proximity assays (SPA), scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of PKIN activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

#### **XVIII. Enhancement/Inhibition of Protein Kinase Activity**

20 Agonists or antagonists of PKIN activation or inhibition may be tested using assays described in section XVII. Agonists cause an increase in PKIN activity and antagonists cause a decrease in PKIN activity.

25 Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polymer ID	Incyte Polynucleotide SEQ ID NO:	Polymer ID	Incyte Polynucleotide ID
058860	1	058860CD1		13		058860CB1
2041716	2	2041716CD1		14		2041716CB1
7472005	3	7472005CD1		15		7472005CB1
7472006	4	7472006CD1		16		7472006CB1
2902460	5	2902460CD1		17		2902460CB1
6383934	6	6383934CD1		18		6383934CB1
3210906	7	3210906CD1		19		3210906CB1
3339024	8	3339024CD1		20		3339024CB1
4436929	9	4436929CD1		21		4436929CB1
5046791	10	5046791CD1		22		5046791CB1
1416174	11	1416174CD1		23		1416174CB1
3244919	12	3244919CD1		24		3244919CB1

Table 2

Polypeptide SEQ ID NO:	Incute Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	058860CD1	g2677788	8.6e-50	Unknown [Sparisoma chrysoporum], related to g4322024, myosin light chain kinase isoform 3B
2	2041716CD1	g1836161	8.3e-253	Ca2+/calmodulin-dependent protein kinase IV kinase isoform [Rattus sp.]
3	7472005CD1	g1750259	0.0	Eph-and Elk-related kinase [Mus musculus]
4	7472006CD1	g404634	3.6e-163	Serine/threonine kinase [Mus musculus]
5	2902460CD1	g396429	4.9e-264	IP3 3-kinase [Rattus norvegicus]
6	6383934CD1	g2738898	5.2e-173	Protein kinase [Mus musculus]
7	3210906CD1	g5616074	0.0	Prostate derived STE20-like kinase PSK [Homo sapiens]
8	33339024CD1	g5295850	4.4e-123	QA79 membrane protein [Homo sapiens] (Falco, M. et al. (1999) J. Exp. Med. 190:793-802)
9	4436929CD1	g1872546	0.0	NIK (Nck Interacting Kinase) [Mus musculus] (Su, Y.C. et al. (1997) EMBO J. 16:1279-1290)
10	5046791CD1	g861314	2.7e-21	Similar to Ser/Thr protein kinase [Caenorhabditis elegans]
11	1416174CD1	g8248287	2.00E-61	sphingosine kinase type 2 isoform [Mus musculus]
12	3244919CD1	g7161864	3.10E-185	serine/threonine protein kinase [Mus musculus]

Table 3

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	058860CD1	466	T422 T5 T12 S19 T31 S46 S83 S168 S179 T194 T331 S351 S365 T422 T52 S163 T299 T312 S402 T451 Y446	N59 N81 N361 N452	Receptor tyrosine kinase: F395-G418 Thiol protease His motif: M116-A126	MOTIFS-BLIMPS-BLOCKS
2	2041716CD1	513	S74 T108 S466 T26 S74 S82 S117 S427 S433 T438 T58 S69 S100 S169 S338 S445	N156	ATP/GTP-binding site motif A (P-loop): G493-S500 Serine/Threonine protein kinase active-site signature: T279-L291 Eukaryotic protein kinase domain: Q145-V417 Tyrosine kinase catalytic domain: Y273-L291, G320-I330, L342-D364 Kinase protein beta: M1-Q127 Protein kinase domain: L130-V408	MOTIFS-BLAST-DOMO-HMMER-PFAM-BLIMPS-PRINTS-BLAST-PRODOM

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	7472005CD1	1012	S56 T104 T117 S129 S136 T155 T219 S225 S374 S577 T615 T805 S817 T843 S856 S857 S897 S926 T941 S177 S196 T242 T489 T494 T531 T674 S848 S908 S948 T997 Y487 Y610 Y756	N340 N407 N432 N718 N841	Eukaryotic protein kinase domain: I635-Y896 Protein kinases ATP-binding region signature: I641-K667 Tyrosine protein kinases specific active-site signature: Y756-V768 Receptor tyrosine kinase class V: C247-E267 (signature 2) E31-H52, D61-P112, K165-V218, P243-E267, C273-P320, V339-V365, C376-S419, S455-K480, G501-T531, P605-G644, P657-M710, L721-M740, L741-A762, A763-P789, G797-W829, E830-V854, F958-Q1001, L34-G380 Tyrosine kinase catalytic domain signature: T713-R726, Y750-V768, I800-I810, S819-N841, C870-F892 Kinase receptor precursor: E31-C204 Ephrin receptor ligand binding domain: E31-C204	MOTIFS HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS BLAST- PRODOM BLAST-DOMO

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	7472006CD1	367	T310 S326 S349 S31 S158 S166 S290 S304		Protein kinases ATP-binding region signature: L18-K41 Serine/Threonine protein kinases active site signature: V132-L144 Eukaryotic protein kinase domain: Y12-M272 Testis specific serine/threonine kinase: M272-T364 Protein kinase domain: L14-I263 Tyrosine kinase catalytic domain signature: M90-K103, Y126-L144, Y197-S219	MOTIFS HMMR-PFAM BLAST- PRODOM BLAST-DOMO BLIMPS- PRINTS
5	2902460CD1	798	S56 S65 T67 T96 S98 T123 S132 S451 T428 S462 S463 Y464 S467 S473 T602 Y603 T634 T715 S18 S69 S116 S179 S292 S324 S386 S440 S499 S515 S531 S616	N317	Signal peptide: M1-A24 Calmodulin-binding domain: DM07435 P42335 210-672: P332-L797 Proline-rich protein: DM01369 B39066 172-256: G274-P330 1-D myoinositol tris-phosphate 3 kinase, EC 2.7.1.127, inositol 1,4,5-tris-phosphate, IP3K, IP3, transferase, kinase, calmodulin- binding: PD138098: G120-S510	SPScan MOTIFS BLAST- PRODOM BLAST-DOMO

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	6383934CD1	358	Y293 T48 S349 S31 S158 S258 S284 T340		Protein kinase ATP-binding domain: L18-K41 Protein kinase ST: I132-L144 Tyrosine kinase catalytic domain signature: M90-K103, Y126-L144, Y197-S219 Eukaryotic protein kinase domain: Y12-L272	MOTIFS PFAM BLIMPS- PRINTS
7	3210906CD1	1049	S306 S9 S111 T214 T346 S370 S375 T671 T701 S806 S853 S894 S1014 S60 S62 S453 T468 S521 T586 T604 T671 S742 T757 T776 T793 T886 S889 S910 T990 Y309	N1042	Protein kinase domain: DM00004 I48609 55-294: L18-R260 Testis specific serine/threonine kinase 2 protein kinase: PD029090: L272-T358 Protein kinase domain: DM00004 JC1446 20-261: V14-I263	BLAST- PRODOM BLAST-DOMO
8	3339024CD1	322	S42 S117 T246 S266 S284 T109 T172 T195 S231 S2316	N17 N87 N94 N112	E618-P777	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	4436929CD1	1212	S77 T187 S259 S608 S873 S9 S17 T59 S112 T124 T222 S264 T319 S324 S326 S548 S567 S604 S627 S680 S739 S740 T746 T747 S764 S778 T989 S1016 S1036 T1050 S1076 S255 S259 T309 T351 T557 T597 S604 S679 S687 S784 T869 S956 S1089 S1190 Y321 Y323 Y467	N33 N546 N624 N776 N1144	Eukaryotic protein kinase domain: F25-I289	HMMER-PFAM
				Protein kinase domain: DM00004   P10676   18-272: L27-P278	BLAST-DO MO	
				CNH domain: Y894-R1192	HMMER-PFAM	
				Protein kinases signatures and profile: W129-T181	PROFILESCAN	
				Protein kinase ST: V149-L161	MOTIFS	
				NIK (Nck Interacting Kinase): PD147187: D496-W908	BLAST- PRODOM	
10	5046791CD1	280	S102 T161 Y162 T92 S209 S243 S102 T161	N155	Protein F55A11.6 C52E4.7, similar to Ser/Thr kinase: PD024191: G11-L130	BLAST- PRODOM
11	1416174CD1	114			Protein chromosome C34C6.5 C4A8.07C 1 sphingosine XII cosmid ORF: PD014044: H8-P97	BLAST- PRODOM
12	3244919CD1	375	S92 S276 T9 T48 T125 S295 T360 Y52	N338	Protein kinase ATP-binding domain: I32-M55 Protein kinase ST: I145-L157	MOTIFS
				Eukaryotic protein kinase domain: F26-Q278	HMMER-PFAM	
				Tyrosine kinase catalytic domain: PR00109: V103-Q116, Y139-L157	BLIMPS- PRINTS	
				Protein kinase domain: DM00004   P54644   122-362: I28-S275 DM08046   P05986   1-397: S3-P305	BLAST-DO MO	

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
13	058860CB1	1859	1-837, 1111-1198	60122573D4 058860R6 (MUSCN0T01) 3011528F6 (MUSCN0T07)	1 370 852	491 1005 1341
14	2041716CB1	3501	1-2773	3016678T6 (MUSCN0T07) 3500745F6 (PROSTUT13) 94454511.v113.gs_3.nt.edit 6063491H1 (BRAEN0T02) 2190612F6 (THYRTUT03) 70168906V1 70164503V1 70168645V1 70167500V1 13833374T6 (BRAITUT08)	1 1 22 715 1072 1392 1840 2056 2541 2688	456 1859 884 1093 1658 1989 2664 2696 3123 3255
15	7472005CB1	3039	1-557, 2741- 3039, 824-1827	95679461.v113.gs_2.edit 95686590.v113.gs_5	1 1	3039 1104
16	7472006CB1	1104	823-1104	70166939V1 6882904J1 (BRAHTDR03)	1 3381 1399	1104 3916 2005
17	2902460CB1	3939	1-1642, 2515-3100, 3766-3939	7117043H1 (BRAHNOE01) 7090661H1 (BRAUTDR03) 6811472J1 (SKIRNOR01) 6882520J1 (BRAHTDR03) 3753286H1 (BRAHDIT04) 7029494H1 (BRAXTDR12)	614 914 2436 1 1692	1253 1492 3034 639 3643 2288
18	6383934CB1	1381	1-359	6911565J1 (PITUDIR01) 7176637H1 (BRSTM0C01) 2695922F6 (UTRSNOT12) 93873504.v113.gs_3.nt 2011686H1 (TESTNOT03) 92821547 6383934H1 (FIBRUNT02) 5281219H1 (TESTNON04)	2169 2766 3114 73 665 972 874 1	2757 3358 3571 1149 858 1381 1176 239

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incute Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
19	3210906CB1	3904	3815-3904, 1-449, 901-1443, 1486-1805, 3039-3432	533823R6 (BRAINNOT03) 1807122F6 (SINTNOT13) 4785178H1 (BRATNOT03) 1439938F6 (THYRNNOT03) 2654018H1 (THYMNOT04) 713861X11 (PROSTUT01) 1416996X310D1 (BRAINOT12) 4326355F6 (TLYMUNT01) 2512189F6 (CONUTUT01) 273994R6 (PANCDDIT03) 860975R6 (BRAITUT03) 273994T6 (PANCDDIT03)	3136 3432 2594 267 1336 1 2383 959 1593 666 2717 1947 446 787 1 1395 1299 3362 555 2886 251 2654 633 1 1427 3153 2312 1919 1085 1 637 586 70495437V1	3683 3904 2860 769 1629 529 2783 1404 2079 1101 3224 2561 1101 1465 629 1987 1588 3925 1107 3425 3925 3136 1149 575 1723 3430 2865 2435 1544 262 1210 1200 880
20	3339024CB1	1987	1-125, 1955-1987, 1461-1493	70774378V1 70772051V1 33339024F6 (SPLNNNOT10) 70775014V1 2986160H1 (CARGDIT01) 1852144T6 (LUNGFFET03) 3136101F6 (SMCCCNNOT01) SCLA03429V1 93327187 CD 2606210F6 (LUNGUTUT07) 2827761F6 (TLYMNNNOT03) SZAU00120V1 3085382H1 (HEAOONOT03) 2956512H1 (KIDNFFET01) SCLA04243V1 1741505R6 (HIPONON01) 2805893F6 (BLADTUT08) 6390331H1 (BONRNOT01) q1512902 260140R6 (HNT2RAT01) 70495437V1	2383 959 1593 666 2717 1947 446 787 1 1395 1299 3362 555 2886 251 2654 633 1 1427 3153 2312 1919 1085 1 637 586 70495437V1	2783 1404 2079 1101 3224 2561 1101 1465 629 1987 1588 3925 1107 3425 3925 3136 1149 575 1723 3430 2865 2435 1544 262 1210 1200 880
21	4436929CB1	3925	1431-2791, 1-956	1852144T6 (LUNGFFET03) 3136101F6 (SMCCCNNOT01) SCLA03429V1 93327187 CD 2606210F6 (LUNGUTUT07) 2827761F6 (TLYMNNNOT03) SZAU00120V1 3085382H1 (HEAOONOT03) 2956512H1 (KIDNFFET01) SCLA04243V1 1741505R6 (HIPONON01) 2805893F6 (BLADTUT08) 6390331H1 (BONRNOT01) q1512902 260140R6 (HNT2RAT01) 70495437V1	1395 1299 3362 555 2886 251 2654 633 1 1427 3153 2312 1919 1085 1 637 586 70495437V1	1987 1588 3925 1107 3425 3925 3136 1149 575 1723 3430 2865 2435 1544 262 1210 1200 880
22	5046791CB1	1210	1-244	6390331H1 (BONRNOT01) q1512902 260140R6 (HNT2RAT01) 70495437V1	1 637 586 70495437V1	262 1210 1200 880

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
23	1416174CB1	1521	1-792, 876-975	3869131H1 (BMARNOT03) 1416174H1 (BRAINOT12) 2169725T6 (ENDCNOT03) 1683338F6 (PROSNOT15) 1284949T6 (COLNNOT16) 3272203F6 (BRAINOT20) 2287966H1 (BRAINON01) 6307341H1 (NERDTDN03) 7177378H1 (BRAXDIC01) 70570341V1 5372702H1 (BRAINOT22) 70568614V1	1 155 933 1030 871 269 898 1429 440 1 1201 1428 677	231 402 1504 1521 1489 898 1640 1134 526 1588 1633 1336
24	3244919CB1	1640	919-1535			

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
13	058860CB1	MUSCN07
14	2041716CB1	BRAXN03
17	2902460CB1	BRAGN02
18	6383934CB1	FIBRINT02
19	3210906CB1	BRAITUT03
20	3339024CB1	THYRN08
21	4436929CB1	ENDCN03
22	5046791CB1	BRABD1R01
23	1416174CB1	CARGDT01
24	3244919CB1	BRAINOT21

Table 6

Library	Vector	Library Description
MUSCN07	PINCY	Library was constructed using 2 micrograms of polyA RNA isolated from muscle tissue removed from the forearm of a 38-year-old Caucasian female during a soft tissue excision. Pathology indicated the surgical margins of re-excision were free of tumor. Pathology for the matched tumor tissue indicated intramuscular hemangioma. Patient history included a normal delivery. Patient medications included melatonin, Valium, and Tylenol PM. Family history included breast cancer in the mother; and benign hypertension, cerebrovascular disease, colon cancer, and type II diabetes in the grandparent(s).
BRAKNOT03	PINCY	Library was constructed using 1.5 micrograms of polyA RNA isolated from sensory-motor cortex tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarcts of the cerebral neocortex. Grossly, the brain regions examined and cranial nerves were unremarkable, showing no evidence of atrophy. No atherosclerosis of the major vessels was noted. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. There were also multiple small microscopic areas of cavitation with surrounding gliosis scattered throughout the cerebral cortex. Special stains with Bielschowsky silver, Kluber-Barrera, and Congo Red revealed no evidence of neurofibrillary tangles or diffuse anorectic amyloid plaques, demyelination, and cerebral amyloid angiopathy, respectively. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly, and an enlarged spleen and liver. Patient medications included simethicone, Lasix, Digoxin, Colace, Zantac, captoril, and Vasotec.

Table 6 cont.

Library	Vector	Library Description
BRAGNON02	PINCY	The library was constructed from a normalized substantia nigra tissue library constructed from 4.2x10 <sup>6</sup> independent clones. Starting RNA was made from RNA isolated from substantia nigra tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloid goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease. The library was normalized in two rounds using conditions adapted from Soares et al., BNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996) :791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
FIBRUNTO2	PINCY	The library was constructed from polyA RNA isolated from an untreated MG-63 cell line derived from an osteosarcoma removed from a 14-year-old Caucasian male
BRABDIR01	PINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, emphysema, and tobacco abuse.
BRAITUT03	PSPORT1	Library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 17-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a grade 4 fibrillary giant and small-cell astrocytoma. Family history included benign hypertension and cerebrovascular disease.
ENDCNOT03	PINCY	Library was constructed using RNA isolated from dermal microvascular endothelial cells removed from a neonatal Caucasian male.

Table 6 cont.

Library	Vector	Library Description
THYRNOT08	pINCY	Library was constructed using RNA isolated from the diseased left thyroid tissue removed from a 13-year-old Caucasian female during a complete thyroidectomy. Pathology indicated lymphocytic thyroiditis. Pathology for the matched tumor tissue indicated grade 1 papillary carcinoma. Multiple lymph nodes from the right, left, and midline section of the neck were negative for tumor. Fragments of the thymus were benign. Fibroadipose tissue was identified in the right inferior and superior parathyroid regions. Multiple lymph nodes (2 of 6) from the right side of the neck contained microscopic foci of metastatic papillary carcinoma. Patient history included attention deficit disorder with hyperactivity. Previous surgeries included an operative procedure on the external ear. Patient medications included Prozac. Family history included chronic obstructive asthma in the mother; alcohol abuse, benign hypertension, and depressive disorder in the grandparent(s); and attention deficit disorder with hyperactivity in the sibling(s).
BRAINOT21	pINCY	Library was constructed using RNA isolated from diseased brain tissue removed from the left frontal lobe of a 46-year-old Caucasian male during a lobectomy. Pathology indicated focal cortical and subcortical scarring of the left frontal lobe, characterized by cavitation and extensive reactive changes, including marked gliosis and hemosiderin deposition, consistent with a history of remote severe head trauma. GFAP was positive in astrocytes. The pattern of reactivity is that of reactive gliosis. Patient history included traumatic intracranial hemorrhage and brain injury with loss of consciousness following head trauma. Family history included cerebrovascular disease, cerebrovascular disease, and atherosclerotic coronary artery disease.
CARDIT01	pINCY	Library was constructed using RNA isolated from diseased cartilage tissue. Patient history included osteoarthritis.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	<i>ESTs</i> : Probability value= 1.0E-8 or less <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, fasta, fastx, fastx, and search.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	<i>ESTs</i> : fasta E value=1.06E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HHMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	<i>PFAM hits</i> : Probability value= 1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Int'l. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

5           a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-12,  
b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12,  
c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and  
10           d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-12.

15           3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

20           5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:13-24.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

25           7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

30           9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

35           b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

- 5        a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24,
- b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24,
- c) a polynucleotide sequence complementary to a),
- d) a polynucleotide sequence complementary to b), and
- 10      e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

15        13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
- 20      b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

25

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- 30      b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

35

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

18. A method for treating a disease or condition associated with decreased expression of 5 functional PKIN, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

10 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and  
b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

15 21. A method for treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 20.

20 22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and  
b) detecting antagonist activity in the sample.

25 23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 23.

30 25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:  
a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and  
35 b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- 5 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in 10 the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method 15 comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of 20 the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at 25 least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the 30 amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

<110> INCYTE GENOMICS, INC.  
YANG, Jumming  
BAUGHN, Mariah R.  
BURFORD, Neil  
AU-YOUNG, Janice  
LU, Dyung Aina M.  
REDDY, Roopa  
YUE, Henry  
YAO, Monique G.  
LAL, Preeti  
KAHN, Farrah A.

<120> HUMAN KINASES

<130> PI-0002 PCT

<140> To Be Assigned  
<141> Herewith

<150> 60/172,066; 60/176,107; 60/176,107; 60/177,731  
<151> 1999-12-23; 2000-01-14; 2000-01-14; 2000-01-21

<160> 24

<170> PERL Program

<210> 1  
<211> 466  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 058860CD1

<400> 1  
Met Glu Asp Gly Thr Pro Asn Glu His Phe Tyr Thr Pro Thr Glu  
1 5 10 15  
Glu Arg Gly Ser Ala Tyr Glu Ile Trp Arg Ser Asp Ser Phe Gly  
20 25 30  
Thr Pro Asn Glu Ala Ile Glu Pro Lys Asp Asn Glu Met Pro Pro  
35 40 45  
Ser Phe Ile Glu Pro Leu Thr Lys Arg Lys Val Tyr Glu Asn Thr  
50 55 60  
Thr Leu Gly Phe Ile Val Glu Val Glu Gly Leu Pro Val Pro Gly  
65 70 75  
Val Lys Trp Tyr Arg Asn Lys Ser Leu Leu Glu Pro Asp Glu Arg  
80 85 90  
Ile Lys Met Glu Arg Val Gly Asn Val Cys Ser Leu Glu Ile Ser  
95 100 105  
Asn Ile Gln Lys Gly Glu Gly Glu Tyr Met Cys His Ala Val  
110 115 120  
Asn Ile Ile Gly Glu Ala Lys Ser Phe Ala Asn Val Asp Ile Met  
125 130 135  
Pro Gln Glu Glu Arg Val Val Ala Leu Pro Pro Pro Val Thr His  
140 145 150  
Gln His Val Met Glu Phe Asp Leu Glu His Thr Thr Ser Ser Arg  
155 160 165  
Thr Pro Ser Pro Gln Glu Ile Val Leu Glu Val Glu Leu Ser Glu  
170 175 180  
Lys Asp Val Lys Glu Phe Glu Lys Gln Val Lys Ile Val Thr Val  
185 190 195  
Pro Glu Phe Thr Pro Asp His Lys Ser Met Ile Val Ser Leu Asp  
200 205 210  
Val Leu Pro Phe Asn Phe Val Asp Pro Asn Met Asp Ser Arg Glu  
215 220 225  
Gly Glu Asp Lys Glu Leu Lys Ile Asp Leu Glu Val Phe Glu Met  
230 235 240

Pro	Pro	Arg	Phe	Ile	Met	Pro	Ile	Cys	Asp	Phe	Lys	Ile	Pro	Glu
					245					250				255
Asn	Ser	Asp	Ala	Val	Phe	Lys	Cys	Ser	Val	Ile	Gly	Ile	Pro	Thr
					260					265				270
Pro	Glu	Val	Lys	Trp	Tyr	Lys	Glu	Tyr	Met	Cys	Ile	Glu	Pro	Asp
					275					280				285
Asn	Ile	Lys	Tyr	Val	Ile	Ser	Glu	Glu	Lys	Gly	Ser	His	Thr	Leu
					290					295				300
Lys	Ile	Arg	Asn	Val	Cys	Leu	Ser	Asp	Ser	Ala	Thr	Tyr	Arg	Cys
					305					310				315
Arg	Ala	Val	Asn	Cys	Val	Gly	Glu	Ala	Ile	Cys	Arg	Gly	Phe	Leu
					320					325				330
Thr	Met	Gly	Asp	Ser	Glu	Ile	Phe	Ala	Val	Ile	Ala	Lys	Lys	Ser
					335					340				345
Lys	Val	Thr	Leu	Ser	Ser	Leu	Met	Glu	Glu	Leu	Val	Leu	Lys	Ser
					350					355				360
Asn	Tyr	Thr	Asp	Ser	Phe	Phe	Glu	Phe	Gln	Val	Val	Glu	Gly	Pro
					365					370				375
Pro	Arg	Phe	Ile	Lys	Gly	Ile	Ser	Asp	Cys	Tyr	Ala	Pro	Ile	Gly
					380					385				390
Thr	Ala	Ala	Tyr	Phe	Gln	Cys	Leu	Val	Arg	Gly	Ser	Pro	Arg	Pro
					395					400				405
Thr	Val	Tyr	Trp	Tyr	Lys	Asp	Gly	Lys	Leu	Val	Gln	Gly	Arg	Arg
					410					415				420
Phe	Thr	Val	Glu	Glu	Ser	Gly	Thr	Gly	Phe	His	Asn	Leu	Phe	Ile
					425					430				435
Thr	Ser	Leu	Val	Lys	Ser	Asp	Glu	Gly	Glu	Tyr	Arg	Cys	Val	Ala
					440					445				450
Thr	Asn	Lys	Ser	Gly	Met	Ala	Glu	Ser	Phe	Ala	Ala	Leu	Thr	Leu
					455					460				465

Thr

<210> 2  
<211> 513  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 2041716CD1

<400> 2														
Met	Glu	Gly	Gly	Pro	Ala	Val	Cys	Cys	Gln	Asp	Pro	Arg	Ala	Glu
1				5					10					15
Leu	Val	Glu	Arg	Val	Ala	Ala	Ile	Asp	Val	Thr	His	Leu	Glu	Glu
					20					25				30
Ala	Asp	Gly	Gly	Pro	Glu	Pro	Thr	Arg	Asn	Gly	Val	Asp	Pro	Pro
					35					40				45
Pro	Arg	Ala	Arg	Ala	Ala	Ser	Val	Ile	Pro	Gly	Ser	Thr	Ser	Arg
					50					55				60
Leu	Leu	Pro	Ala	Arg	Pro	Ser	Leu	Ser	Ala	Arg	Lys	Leu	Ser	Leu
					65					70				75
Gln	Glu	Arg	Pro	Ala	Gly	Ser	Tyr	Leu	Glu	Ala	Gln	Ala	Gly	Pro
					80					85				90
Tyr	Ala	Thr	Gly	Pro	Ala	Ser	His	Ile	Ser	Pro	Arg	Ala	Trp	Arg
					95					100				105
Arg	Pro	Thr	Ile	Glu	Ser	His	His	Val	Ala	Ile	Ser	Asp	Ala	Glu
					110					115				120
Asp	Cys	Val	Gln	Leu	Asn	Gln	Tyr	Lys	Leu	Gln	Ser	Glu	Ile	Gly
					125					130				135
Lys	Val	Gly	Leu	Thr	Asp	Ala	Tyr	Leu	Gln	Gly	Ala	Tyr	Gly	Val
					140					145				150
Val	Arg	Leu	Ala	Tyr	Asn	Glu	Ser	Glu	Asp	Arg	His	Tyr	Ala	Met
					155					160				165
Lys	Val	Leu	Ser	Lys	Lys	Lys	Leu	Leu	Lys	Gln	Tyr	Gly	Phe	Pro
					170					175				180

Arg Arg Pro Pro Pro Arg Gly Ser Gln Ala Ala Gln Gly Gly Pro  
 185 190 195  
 Ala Lys Gln Leu Leu Pro Leu Glu Arg Val Tyr Gln Glu Ile Ala  
 200 205 210  
 Ile Leu Lys Lys Leu Asp His Val Asn Val Val Lys Leu Ile Glu  
 215 220 225  
 Val Leu Asp Asp Pro Ala Glu Asp Asn Leu Tyr Leu Val Asp Leu  
 230 235 240  
 Leu Arg Lys Gly Pro Val Met Glu Val Pro Cys Asp Lys Pro Phe  
 245 250 255  
 Ser Glu Glu Gln Ala Arg Leu Tyr Leu Arg Asp Val Ile Leu Gly  
 260 265 270  
 Leu Glu Tyr Leu His Cys Gln Lys Ile Val His Arg Asp Ile Lys  
 275 280 285  
 Pro Ser Asn Leu Leu Leu Gly Asp Asp Gly His Val Lys Ile Ala  
 290 295 300  
 Asp Phe Gly Val Ser Asn Gln Phe Glu Gly Asn Asp Ala Gln Leu  
 305 310 315  
 Ser Ser Thr Ala Gly Thr Pro Ala Phe Met Ala Pro Glu Ala Ile  
 320 325 330  
 Ser Asp Ser Gly Gln Ser Phe Ser Gly Lys Ala Leu Asp Val Trp  
 335 340 345  
 Ala Thr Gly Val Thr Leu Tyr Cys Phe Val Tyr Gly Lys Cys Pro  
 350 355 360  
 Phe Ile Asp Asp Phe Ile Leu Ala Leu His Arg Lys Ile Lys Asn  
 365 370 375  
 Glu Pro Val Val Phe Pro Glu Glu Pro Glu Ile Ser Glu Glu Leu  
 380 385 390  
 Lys Asp Leu Ile Leu Lys Met Leu Asp Lys Asn Pro Glu Thr Arg  
 395 400 405  
 Ile Gly Val Pro Asp Ile Lys Leu His Pro Trp Val Thr Lys Asn  
 410 415 420  
 Gly Glu Glu Pro Leu Pro Ser Glu Glu Glu His Cys Ser Val Val  
 425 430 435  
 Glu Val Thr Glu Glu Glu Val Lys Asn Ser Val Arg Leu Ile Pro  
 440 445 450  
 Ser Trp Thr Thr Val Ile Leu Val Lys Ser Met Leu Arg Lys Arg  
 455 460 465  
 Ser Phe Gly Asn Pro Phe Glu Pro Gln Ala Arg Arg Glu Glu Arg  
 470 475 480  
 Ser Met Ser Ala Pro Gly Asn Leu Leu Val Lys Glu Gly Phe Gly  
 485 490 495  
 Glu Gly Gly Lys Ser Pro Glu Leu Pro Gly Val Gln Glu Asp Glu  
 500 505 510  
 Ala Ala Ser

<210> 3  
 <211> 1012  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7472005CD1

<400> 3  
 Met Ala Pro Ala Arg Gly Arg Leu Pro Pro Ala Leu Trp Val Val  
 1 5 10 15  
 Thr Ala Ala Ala Ala Ala Ala Thr Cys Val Ser Ala Ala Arg Gly  
 20 25 30  
 Glu Val Asn Leu Leu Asp Thr Ser Thr Ile His Gly Asp Trp Gly  
 35 40 45  
 Trp Leu Thr Tyr Pro Ala His Gly Trp Asp Ser Ile Asn Glu Val  
 50 55 60  
 Asp Glu Ser Phe Gln Pro Ile His Thr Tyr Gln Val Cys Asn Val  
 65 70 75

Met Ser Pro Asn Gln Asn Asn Trp Leu Arg Thr Ser Trp Val Pro  
                   80                 85                 90  
 Arg Asp Gly Ala Arg Arg Val Tyr Ala Glu Ile Lys Phe Thr Leu  
                   95                 100                 105  
 Arg Asp Cys Asn Ser Met Pro Gly Val Leu Gly Thr Cys Lys Glu  
                  110                 115                 120  
 Thr Phe Asn Leu Tyr Tyr Leu Glu Ser Asp Arg Asp Leu Gly Ala  
                  125                 130                 135  
 Ser Thr Gln Glu Ser Gln Phe Leu Lys Ile Asp Thr Ile Ala Ala  
                  140                 145                 150  
 Asp Glu Ser Phe Thr Gly Ala Asp Leu Gly Val Arg Arg Leu Lys  
                  155                 160                 165  
 Leu Asn Thr Glu Val Arg Ser Val Gly Pro Leu Ser Lys Arg Gly  
                  170                 175                 180  
 Phe Tyr Leu Ala Phe Gln Asp Ile Gly Ala Cys Leu Ala Ile Leu  
                  185                 190                 195  
 Ser Leu Arg Ile Tyr Tyr Lys Lys Cys Pro Ala Met Val Arg Asn  
                  200                 205                 210  
 Leu Ala Ala Phe Ser Glu Ala Val Thr Gly Ala Asp Ser Ser Ser  
                  215                 220                 225  
 Leu Val Glu Val Arg Gly Gln Cys Val Arg His Ser Glu Glu Arg  
                  230                 235                 240  
 Asp Thr Pro Lys Met Tyr Cys Ser Ala Glu Gly Glu Trp Leu Val  
                  245                 250                 255  
 Pro Ile Gly Lys Cys Val Cys Ser Ala Gly Tyr Glu Glu Arg Arg  
                  260                 265                 270  
 Asp Ala Cys Val Ala Cys Glu Leu Gly Phe Tyr Lys Ser Ala Pro  
                  275                 280                 285  
 Gly Asp Gln Leu Cys Ala Arg Cys Pro Pro His Ser His Ser Ala  
                  290                 295                 300  
 Ala Pro Ala Ala Gln Ala Cys His Cys Asp Leu Ser Tyr Tyr Arg  
                  305                 310                 315  
 Ala Ala Leu Asp Pro Pro Ser Ser Ala Cys Thr Arg Pro Pro Ser  
                  320                 325                 330  
 Ala Pro Val Asn Leu Ile Ser Ser Val Asn Gly Thr Ser Val Thr  
                  335                 340                 345  
 Leu Glu Trp Ala Pro Pro Leu Asp Pro Gly Gly Arg Ser Asp Ile  
                  350                 355                 360  
 Thr Tyr Asn Ala Val Cys Arg Arg Cys Pro Trp Ala Leu Ser Arg  
                  365                 370                 375  
 Cys Glu Ala Cys Gly Ser Gly Thr Arg Phe Val Pro Gln Gln Thr  
                  380                 385                 390  
 Ser Leu Val Gln Ala Ser Leu Leu Val Ala Asn Leu Leu Ala His  
                  395                 400                 405  
 Met Asn Tyr Ser Phe Trp Ile Glu Ala Val Asn Gly Val Ser Asp  
                  410                 415                 420  
 Leu Ser Pro Glu Pro Arg Arg Ala Ala Val Val Asn Ile Thr Thr  
                  425                 430                 435  
 Asn Gln Ala Ala Pro Ser Gln Val Val Val Ile Arg Gln Glu Arg  
                  440                 445                 450  
 Ala Gly Gln Thr Ser Val Ser Leu Leu Trp Gln Glu Pro Glu Gln  
                  455                 460                 465  
 Pro Asn Gly Ile Ile Leu Glu Tyr Glu Ile Lys Tyr Tyr Glu Lys  
                  470                 475                 480  
 Asp Lys Glu Met Gln Ser Tyr Ser Thr Leu Lys Ala Val Thr Thr  
                  485                 490                 495  
 Arg Ala Thr Val Ser Gly Leu Lys Pro Gly Thr Arg Tyr Val Phe  
                  500                 505                 510  
 Gln Val Arg Ala Arg Thr Ser Ala Gly Cys Gly Arg Phe Ser Gln  
                  515                 520                 525  
 Ala Met Glu Val Glu Thr Gly Lys Pro Arg Pro Arg Tyr Asp Thr  
                  530                 535                 540  
 Arg Thr Ile Val Trp Ile Cys Leu Thr Leu Ile Thr Gly Leu Val  
                  545                 550                 555  
 Val Leu Leu Leu Leu Leu Ile Cys Lys Lys Arg His Cys Gly Tyr  
                  560                 565                 570  
 Ser Lys Ala Phe Gln Asp Ser Asp Glu Glu Lys Met His Tyr Gln

Asn	Gly	Gln	Ala	Pro	Pro	Pro	Val	Phe	Leu	Pro	Leu	His	His	Pro
				575					580					585
				590					595					600
Pro	Gly	Lys	Leu	Pro	Glu	Pro	Gln	Phe	Tyr	Ala	Glu	Pro	His	Thr
				605					610					615
Tyr	Glu	Glu	Pro	Gly	Arg	Ala	Gly	Arg	Ser	Phe	Thr	Arg	Glu	Ile
				620					625					630
Glu	Ala	Ser	Arg	Ile	His	Ile	Glu	Lys	Ile	Ile	Gly	Ser	Gly	Asp
				635					640					645
Ser	Gly	Glu	Val	Cys	Tyr	Gly	Arg	Leu	Arg	Val	Pro	Gly	Gln	Arg
				650					655					660
Asp	Val	Pro	Val	Ala	Ile	Lys	Ala	Leu	Lys	Ala	Gly	Tyr	Thr	Glu
				665					670					675
Arg	Gln	Arg	Arg	Asp	Phe	Leu	Ser	Glu	Ala	Ser	Ile	Met	Gly	Gln
				680					685					690
Phe	Asp	His	Pro	Asn	Ile	Ile	Arg	Leu	Glu	Gly	Val	Val	Thr	Arg
				695					700					705
Gly	Arg	Leu	Ala	Met	Ile	Val	Thr	Glu	Tyr	Met	Glu	Asn	Gly	Ser
				710					715					720
Leu	Asp	Thr	Phe	Leu	Arg	Thr	His	Asp	Gly	Gln	Phe	Thr	Ile	Met
				725					730					735
Gln	Leu	Val	Gly	Met	Leu	Arg	Gly	Val	Gly	Ala	Gly	Met	Arg	Tyr
				740					745					750
Leu	Ser	Asp	Leu	Gly	Tyr	Val	His	Arg	Asp	Leu	Ala	Ala	Arg	Asn
				755					760					765
Val	Leu	Val	Asp	Ser	Asn	Leu	Val	Cys	Lys	Val	Ser	Asp	Phe	Gly
				770					775					780
Leu	Ser	Arg	Val	Leu	Glu	Asp	Asp	Pro	Asp	Ala	Ala	Tyr	Thr	Thr
				785					790					795
Thr	Gly	Gly	Lys	Ile	Pro	Ile	Arg	Trp	Thr	Ala	Pro	Glu	Ala	Ile
				800					805					810
Ala	Phe	Arg	Thr	Phe	Ser	Ser	Ala	Ser	Asp	Val	Trp	Ser	Phe	Gly
				815					820					825
Val	Val	Met	Trp	Glu	Val	Leu	Ala	Tyr	Gly	Glu	Arg	Pro	Tyr	Trp
				830					835					840
Asn	Met	Thr	Asn	Arg	Asp	Val	Ser	Ala	Lys	Pro	Trp	Gln	Val	Ile
				845					850					855
Ser	Ser	Val	Glu	Glu	Gly	Tyr	Arg	Leu	Pro	Ala	Pro	Met	Gly	Cys
				860					865					870
Pro	His	Ala	Leu	His	Gln	Leu	Met	Leu	Asp	Cys	Trp	His	Lys	Asp
				875					880					885
Arg	Ala	Gln	Arg	Pro	Arg	Phe	Ser	Gln	Ile	Val	Ser	Val	Leu	Asp
				890					895					900
Ala	Leu	Ile	Arg	Ser	Pro	Glu	Ser	Leu	Arg	Ala	Thr	Ala	Thr	Val
				905					910					915
Ser	Arg	Cys	Pro	Pro	Pro	Ala	Phe	Val	Arg	Ser	Cys	Phe	Asp	Leu
				920					925					930
Arg	Gly	Gly	Ser	Gly	Gly	Gly	Gly	Gly	Leu	Thr	Val	Gly	Asp	Trp
				935					940					945
Leu	Asp	Ser	Ile	Arg	Met	Gly	Arg	Tyr	Arg	Asp	His	Phe	Ala	Ala
				950					955					960
Gly	Gly	Tyr	Ser	Ser	Leu	Gly	Met	Val	Leu	Arg	Met	Asn	Ala	Gln
				965					970					975
Asp	Val	Arg	Ala	Leu	Gly	Ile	Thr	Leu	Met	Gly	His	Gln	Lys	Lys
				980					985					990
Ile	Leu	Gly	Ser	Ile	Gln	Thr	Met	Arg	Ala	Gln	Leu	Thr	Ser	Thr
				995					1000					1005
Gln	Gly	Pro	Arg	Arg	His	Leu								
				1010										

<210> 4  
 <211> 367  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature

<223> Incyte ID No: 7472006CD1

<400> 4

Met Asp Asp Ala Ala Val Leu Lys Arg Arg Gly Tyr Leu Leu Gly  
 1 5 10 15  
 Ile Asn Leu Gly Glu Gly Ser Tyr Ala Lys Val Lys Ser Ala Tyr  
 20 25 30  
 Ser Glu Arg Leu Lys Phe Asn Val Ala Ile Lys Ile Ile Asp Arg  
 35 40 45  
 Lys Lys Ala Pro Ala Asp Phe Leu Glu Lys Phe Leu Pro Arg Glu  
 50 55 60  
 Ile Glu Ile Leu Ala Met Leu Asn His Cys Ser Ile Ile Lys Thr  
 65 70 75  
 Tyr Glu Ile Phe Glu Thr Ser His Gly Lys Val Tyr Ile Val Met  
 80 85 90  
 Glu Leu Ala Val Gln Gly Asp Leu Leu Glu Leu Ile Lys Thr Arg  
 95 100 105  
 Gly Ala Leu His Glu Asp Glu Ala Arg Lys Lys Phe His Gln Leu  
 110 115 120  
 Ser Leu Ala Ile Lys Tyr Cys His Asp Leu Asp Val Val His Arg  
 125 130 135  
 Asp Leu Lys Cys Asp Asn Leu Leu Leu Asp Lys Asp Phe Asn Ile  
 140 145 150  
 Lys Leu Ser Asp Phe Ser Phe Ser Lys Arg Cys Leu Arg Asp Asp  
 155 160 165  
 Ser Gly Arg Met Ala Leu Ser Lys Thr Phe Cys Gly Ser Pro Ala  
 170 175 180  
 Tyr Ala Ala Pro Glu Val Leu Gln Gly Ile Pro Tyr Gln Pro Lys  
 185 190 195  
 Val Tyr Asp Ile Trp Ser Leu Gly Val Ile Leu Tyr Ile Met Val  
 200 205 210  
 Cys Gly Ser Met Pro Tyr Asp Asp Ser Asn Ile Lys Lys Met Leu  
 215 220 225  
 Arg Ile Gln Lys Glu His Arg Val Asn Phe Pro Arg Ser Lys His  
 230 235 240  
 Leu Thr Gly Glu Cys Lys Asp Leu Ile Tyr His Met Leu Gln Pro  
 245 250 255  
 Asp Val Asn Arg Arg Leu His Ile Asp Glu Ile Leu Ser His Cys  
 260 265 270  
 Trp Met Gln Pro Lys Ala Arg Gly Ser Pro Ser Val Ala Ile Asn  
 275 280 285  
 Lys Glu Gly Glu Ser Ser Arg Gly Thr Glu Pro Leu Trp Thr Pro  
 290 295 300  
 Glu Pro Gly Ser Asp Lys Lys Ser Ala Thr Lys Leu Glu Pro Glu  
 305 310 315  
 Gly Glu Ala Gln Pro Gln Ala Gln Pro Glu Thr Lys Pro Glu Gly  
 320 325 330  
 Thr Ala Met Gln Met Ser Arg Gln Ser Glu Ile Leu Gly Phe Pro  
 335 340 345  
 Ser Lys Pro Ser Thr Met Glu Thr Glu Glu Gly Pro Pro Gln Gln  
 350 355 360  
 Pro Pro Glu Thr Arg Ala Gln  
 365

<210> 5

<211> 798

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2902460CD1

<400> 5

Met Phe Glu Ala His Ile Gln Ala Gln Ser Ser Ala Ile Gln Ala  
 1 5 10 15  
 Pro Arg Ser Pro Arg Leu Gly Arg Ala Arg Ser Pro Ser Pro Cys

20	25	30
Pro Phe Arg Ser Ser	Ser Gln Pro Pro Gly	Arg Val Leu Val Gln
35	40	45
Gly Ala Arg Ser Glu	Glu Arg Arg Thr Lys	Ser Trp Gly Glu Gln
50	55	60
Cys Pro Glu Thr Ser	Gly Thr Asp Ser Gly	Arg Lys Gly Gly Pro
65	70	75
Ser Leu Cys Ser Ser	Gln Val Lys Lys Gly	Met Pro Pro Leu Pro
80	85	90
Gly Arg Ala Ala Pro	Thr Gly Ser Glu Ala	Gln Gly Pro Ser Ala
95	100	105
Phe Val Arg Met Glu	Lys Gly Ile Pro Ala	Ser Pro Arg Cys Gly
110	115	120
Ser Pro Thr Ala Met	Glu Ile Asp Lys Arg	Gly Ser Pro Thr Pro
125	130	135
Gly Thr Arg Ser Cys	Leu Ala Pro Ser Leu	Gly Leu Phe Gly Ala
140	145	150
Ser Leu Thr Met Ala	Thr Glu Val Ala Ala	Arg Val Thr Ser Thr
155	160	165
Gly Pro His Arg Pro	Gln Asp Leu Ala Leu	Thr Glu Pro Ser Gly
170	175	180
Arg Ala Arg Glu Leu	Glu Asp Leu Gln Pro	Pro Pro Glu Ala Leu
185	190	195
Glu Arg Gln Gly Gln	Phe Leu Gly Ser Glu	Thr Ser Pro Ala Pro
200	205	210
Glu Arg Gly Gly Pro	Arg Asp Gly Glu Pro	Pro Gly Lys Met Gly
215	220	225
Lys Gly Tyr Leu Pro	Cys Gly Met Pro Gly	Ser Gly Glu Pro Glu
230	235	240
Val Gly Lys Arg Pro	Glu Glu Thr Thr Val	Ser Val Gln Ser Ala
245	250	255
Glu Ser Ser Asp Ala	Leu Ser Trp Ser Arg	Leu Pro Arg Ala Leu
260	265	270
Ala Ser Val Gly Pro	Glu Glu Ala Arg Ser	Gly Ala Pro Val Gly
275	280	285
Gly Gly Arg Trp Gln	Leu Ser Asp Arg Val	Glu Gly Gly Ser Pro
290	295	300
Thr Leu Gly Leu Leu	Gly Gly Ser Pro Ser	Ala Gln Pro Gly Thr
305	310	315
Gly Asn Val Glu Ala	Gly Ile Pro Ser Gly	Arg Met Leu Glu Pro
320	325	330
Leu Pro Cys Trp Asp	Ala Ala Lys Asp Leu	Lys Glu Pro Gln Cys
335	340	345
Pro Pro Gly Asp Arg	Val Gly Val Gln Pro	Gly Asn Ser Arg Val
350	355	360
Trp Gln Gly Thr Met	Glu Lys Ala Gly Leu	Ala Trp Thr Arg Gly
365	370	375
Thr Gly Val Gln Ser	Glu Gly Thr Trp Glu	Ser Gln Arg Gln Asp
380	385	390
Ser Asp Ala Leu Pro	Ser Pro Glu Leu Leu	Pro Gln Asp Gln Asp
395	400	405
Lys Pro Phe Leu Arg	Lys Ala Cys Ser Pro	Ser Asn Ile Pro Ala
410	415	420
Val Ile Ile Thr Asp	Met Gly Thr Gln Glu	Asp Gly Ala Leu Glu
425	430	435
Glu Thr Gln Gly Ser	Pro Arg Gly Asn Leu	Pro Leu Arg Lys Leu
440	445	450
Ser Ser Ser Ser Ala	Ser Ser Thr Gly Phe	Ser Ser Ser Tyr Glu
455	460	465
Asp Ser Glu Glu Asp	Ile Ser Ser Asp Pro	Glu Arg Thr Leu Asp
470	475	480
Pro Asn Ser Ala Phe	Leu His Thr Leu Asp	Gln Gln Lys Pro Arg
485	490	495
Val Ser Lys Ser Trp	Arg Lys Ile Lys Asn	Met Val His Trp Ser
500	505	510
Pro Phe Val Met Ser	Phe Lys Lys Tyr	Pro Trp Ile Gln Leu
515	520	525

Ala Gly His Ala Gly Ser Phe Lys Ala Ala Ala Asn Gly Arg Ile  
 530 535 540  
 Leu Lys Lys His Cys Glu Ser Glu Gln Arg Cys Leu Asp Arg Leu  
 545 550 555  
 Met Val Asp Val Leu Arg Pro Phe Val Pro Ala Tyr His Gly Asp  
 560 565 570  
 Val Val Lys Asp Gly Glu Arg Tyr Asn Gln Met Asp Asp Leu Leu  
 575 580 585  
 Ala Asp Phe Asp Ser Pro Cys Val Met Asp Cys Lys Met Gly Ile  
 590 595 600  
 Arg Thr Tyr Leu Glu Glu Glu Leu Thr Lys Ala Arg Lys Lys Pro  
 605 610 615  
 Ser Leu Arg Lys Asp Met Tyr Gln Lys Met Ile Glu Val Asp Pro  
 620 625 630  
 Glu Ala Pro Thr Glu Glu Glu Lys Ala Gln Arg Ala Val Thr Lys  
 635 640 645  
 Pro Arg Tyr Met Gln Trp Arg Glu Thr Ile Ser Ser Thr Ala Thr  
 650 655 660  
 Leu Gly Phe Arg Ile Glu Gly Ile Lys Lys Glu Asp Gly Thr Val  
 665 670 675  
 Asn Arg Asp Phe Lys Lys Thr Lys Thr Arg Glu Gln Val Thr Glu  
 680 685 690  
 Ala Phe Arg Glu Phe Thr Lys Gly Asn His Asn Ile Leu Ile Ala  
 695 700 705  
 Tyr Arg Asp Arg Leu Lys Ala Ile Arg Thr Thr Leu Glu Val Ser  
 710 715 720  
 Pro Phe Phe Lys Cys His Glu Val Ile Gly Ser Ser Leu Leu Phe  
 725 730 735  
 Ile His Asp Lys Lys Glu Gln Ala Lys Val Trp Met Ile Asp Phe  
 740 745 750  
 Gly Lys Thr Thr Pro Leu Pro Glu Gly Gln Thr Leu Gln His Asp  
 755 760 765  
 Val Pro Trp Gln Glu Gly Asn Arg Glu Asp Gly Tyr Leu Ser Gly  
 770 775 780  
 Leu Asn Asn Leu Val Asp Ile Leu Thr Glu Met Ser Gln Asp Ala  
 785 790 795  
 Pro Leu Ala

<210> 6  
 <211> 358  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 6383934CD1

<400> 6  
 Met Asp Asp Ala Thr Val Leu Arg Lys Lys Gly Tyr Ile Val Gly  
 1 5 10 15  
 Ile Asn Leu Gly Lys Gly Ser Tyr Ala Lys Val Lys Ser Ala Tyr  
 20 25 30  
 Ser Glu Arg Leu Lys Phe Asn Val Ala Val Lys Ile Ile Asp Arg  
 35 40 45  
 Lys Lys Thr Pro Thr Asp Phe Val Glu Arg Phe Leu Pro Arg Glu  
 50 55 60  
 Met Asp Ile Leu Ala Thr Val Asn His Gly Ser Ile Ile Lys Thr  
 65 70 75  
 Tyr Glu Ile Phe Glu Thr Ser Asp Gly Arg Ile Tyr Ile Ile Met  
 80 85 90  
 Glu Leu Gly Val Gln Gly Asp Leu Leu Glu Phe Ile Lys Cys Gln  
 95 100 105  
 Gly Ala Leu His Glu Asp Val Ala Arg Lys Met Phe Arg Gln Leu  
 110 115 120  
 Ser Ser Ala Val Lys Tyr Cys His Asp Leu Asp Ile Val His Arg  
 125 130 135

Asp	Leu	Lys	Cys	Glu	Asn	Leu	Leu	Leu	Asp	Lys	Asp	Phe	Asn	Ile
				140					145					150
Lys	Leu	Ser	Asp	Phe	Gly	Phe	Ser	Lys	Arg	Cys	Leu	Arg	Asp	Ser
				155					160					165
Asn	Gly	Arg	Ile	Ile	Leu	Ser	Lys	Thr	Phe	Cys	Gly	Ser	Ala	Ala
				170					175					180
Tyr	Ala	Ala	Pro	Glu	Val	Leu	Gln	Ser	Ile	Pro	Tyr	Gln	Pro	Lys
				185					190					195
Val	Tyr	Asp	Ile	Trp	Ser	Leu	Gly	Val	Ile	Leu	Tyr	Ile	Met	Val
				200					205					210
Cys	Gly	Ser	Met	Pro	Tyr	Asp	Asp	Ser	Asp	Ile	Lys	Lys	Met	Leu
				215					220					225
Arg	Ile	Gln	Lys	Glu	His	Arg	Val	Asn	Phe	Pro	Arg	Ser	Lys	His
				230					235					240
Leu	Thr	Cys	Glu	Cys	Lys	Asp	Leu	Ile	Tyr	His	Met	Leu	Gln	Pro
				245					250					255
Asp	Val	Ser	Gln	Arg	Leu	His	Ile	Asp	Glu	Ile	Leu	Ser	His	Ser
				260					265					270
Trp	Leu	Gln	Pro	Pro	Lys	Pro	Lys	Ala	Thr	Ser	Ser	Ala	Ser	Phe
				275					280					285
Lys	Arg	Glu	Gly	Glu	Gly	Lys	Tyr	Arg	Ala	Glu	Cys	Lys	Leu	Asp
				290					295					300
Thr	Lys	Thr	Gly	Leu	Arg	Pro	Asp	His	Arg	Pro	Asp	His	Lys	Leu
				305					310					315
Gly	Ala	Lys	Thr	Gln	His	Arg	Leu	Leu	Val	Val	Pro	Glu	Asn	Glu
				320					325					330
Asn	Arg	Met	Glu	Asp	Arg	Leu	Ala	Glu	Thr	Ser	Arg	Ala	Lys	Asp
				335					340					345
His	His	Ile	Ser	Gly	Ala	Glu	Val	Gly	Lys	Ala	Ser	Thr		
				350					355					

<210> 7  
<211> 1049

<212> PRT

<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3210906CD1

<400> 7															
Met	Pro	Ala	Gly	Gly	Arg	Ala	Gly	Ser	Leu	Lys	Asp	Pro	Asp	Val	
1				5					10					15	
Ala	Glu	Leu	Phe	Phe	Lys	Asp	Asp	Pro	Glu	Lys	Leu	Phe	Ser	Asp	
				20					25					30	
Leu	Arg	Glu	Ile	Gly	His	Gly	Ser	Phe	Gly	Ala	Val	Tyr	Phe	Ala	
				35					40					45	
Arg	Asp	Val	Arg	Asn	Ser	Glu	Val	Val	Ala	Ile	Lys	Lys	Met	Ser	
				50					55					60	
Tyr	Ser	Gly	Lys	Gln	Ser	Asn	Glu	Lys	Trp	Gln	Asp	Ile	Ile	Lys	
				65					70					75	
Glu	Val	Arg	Phe	Leu	Gln	Lys	Leu	Arg	His	Pro	Asn	Thr	Ile	Gln	
				80					85					90	
Tyr	Arg	Gly	Cys	Tyr	Leu	Arg	Glu	His	Thr	Ala	Trp	Leu	Val	Met	
				95					100					105	
Glu	Tyr	Cys	Leu	Gly	Ser	Thr	Ser	Asp	Leu	Leu	Glu	Val	His	Lys	
				110					115					120	
Lys	Pro	Leu	Gln	Glu	Val	Glu	Ile	Ala	Ala	Val	Thr	His	Gly	Ala	
				125					130					135	
Leu	Gln	Gly	Leu	Ala	Tyr	Leu	His	Ser	His	Asn	Met	Ile	His	Arg	
				140					145					150	
Asp	Val	Lys	Ala	Gly	Asn	Ile	Leu	Leu	Ser	Glu	Pro	Gly	Leu	Val	
				155					160					165	
Lys	Leu	Gly	Asp	Phe	Gly	Ser	Ala	Ser	Ile	Met	Ala	Pro	Ala	Asn	
				170					175					180	
Ser	Phe	Val	Gly	Thr	Pro	Tyr	Trp	Met	Ala	Pro	Glu	Val	Ile	Leu	
				185					190					195	

Ala Met Asp Glu Gly Gln Tyr Asp Gly Lys Val Asp Val Trp Ser  
 200 205 210  
 Leu Gly Ile Thr Cys Ile Glu Leu Ala Glu Arg Lys Pro Pro Leu  
 215 220 225  
 Phe Asn Met Asn Ala Met Ser Ala Leu Tyr His Ile Ala Gln Asn  
 230 235 240  
 Glu Ser Pro Val Leu Gln Ser Gly His Trp Ser Glu Tyr Phe Arg  
 245 250 255  
 Asn Phe Val Asp Ser Cys Leu Gln Lys Ile Pro Gln Asp Arg Pro  
 260 265 270  
 Thr Ser Glu Val Leu Leu Lys His Arg Phe Val Leu Arg Glu Arg  
 275 280 285  
 Pro Pro Thr Val Ile Met Asp Leu Ile Gln Arg Thr Lys Asp Ala  
 290 295 300  
 Val Arg Glu Leu Asp Ser Leu Gln Tyr Arg Lys Met Lys Lys Ile  
 305 310 315  
 Leu Phe Gln Glu Ala Pro Asn Gly Pro Gly Ala Glu Ala Pro Glu  
 320 325 330  
 Glu Glu Glu Ala Glu Pro Tyr Met His Leu Ala Gly Thr Leu  
 335 340 345  
 Thr Ser Leu Glu Ser Ser His Ser Val Pro Ser Met Ser Ile Ser  
 350 355 360  
 Ala Ser Ser Gln Ser Ser Ser Val Asn Ser Leu Ala Asp Ala Ser  
 365 370 375  
 Asp Asn Glu  
 380 385 390  
 Glu Glu Gly Pro Glu Ala Arg Glu Met Ala Met Met Gln Glu Gly  
 395 400 405  
 Glu His Thr Val Thr Ser His Ser Ser Ile Ile His Arg Leu Pro  
 410 415 420  
 Gly Ser Asp Asn Leu Tyr Asp Asp Pro Tyr Gln Pro Glu Ile Thr  
 425 430 435  
 Pro Ser Pro Leu Gln Pro Pro Ala Ala Pro Ala Pro Thr Ser Thr  
 440 445 450  
 Thr Ser Ser Ala Arg Arg Arg Ala Tyr Cys Arg Asn Arg Asp His  
 455 460 465  
 Phe Ala Thr Ile Arg Thr Ala Ser Leu Val Ser Arg Gln Ile Gln  
 470 475 480  
 Glu His Glu Gln Asp Ser Ala Leu Arg Glu Gln Leu Ser Gly Tyr  
 485 490 495  
 Lys Arg Met Arg Arg Gln His Gln Lys Gln Leu Ala Leu Glu  
 500 505 510  
 Ser Arg Leu Arg Gly Glu Arg Glu Glu His Ser Ala Arg Leu Gln  
 515 520 525  
 Arg Glu Leu Glu Ala Gln Arg Ala Gly Phe Gly Ala Glu Ala Glu  
 530 535 540  
 Lys Leu Ala Arg Arg His Gln Ala Ile Gly Glu Lys Glu Ala Arg  
 545 550 555  
 Ala Ala Gln Ala Glu Glu Arg Lys Phe Gln Gln His Ile Leu Gly  
 560 565 570  
 Gln Gln Lys Lys Glu Leu Ala Ala Leu Leu Glu Ala Gln Lys Arg  
 575 580 585  
 Thr Tyr Lys Leu Arg Lys Glu Gln Leu Lys Glu Glu Leu Gln Glu  
 590 595 600  
 Asn Pro Ser Thr Pro Lys Arg Glu Lys Ala Glu Trp Leu Leu Arg  
 605 610 615  
 Gln Lys Glu Gln Leu Gln Gln Cys Gln Ala Glu Glu Glu Ala Gly  
 620 625 630  
 Leu Leu Arg Arg Gln Arg Gln Tyr Phe Glu Leu Gln Cys Arg Gln  
 635 640 645  
 Tyr Lys Arg Lys Met Leu Leu Ala Arg His Ser Leu Asp Gln Asp  
 650 655 660  
 Leu Leu Arg Glu Asp Leu Asn Lys Lys Gln Thr Gln Lys Asp Leu  
 665 670 675  
 Glu Cys Ala Leu Leu Leu Arg Gln His Glu Ala Thr Arg Glu Leu  
 680 685 690  
 Glu Leu Arg Gln Leu Gln Ala Val Gln Arg Thr Arg Ala Glu Leu

695	700	705
Thr Arg Leu Gln His Gln Thr Glu Leu Gly Asn Gln Leu Glu		Tyr
710	715	720
Asn Lys Arg Arg Glu Gln Glu Leu Arg Gln Lys His Ala Ala Gln		
725	730	735
Val Arg Gln Gln Pro Lys Ser Leu Lys Ser Lys Glu Leu Gln Ile		
740	745	750
Lys Lys Gln Phe Gln Glu Thr Cys Lys Ile Gln Thr Arg Gln Tyr		
755	760	765
Lys Ala Leu Arg Ala His Leu Leu Glu Thr Thr Pro Lys Ala Gln		
770	775	780
His Lys Ser Leu Leu Lys Arg Leu Lys Glu Glu Gln Thr Arg Lys		
785	790	795
Leu Ala Ile Leu Ala Glu Gln Tyr Asp Gln Ser Ile Ser Glu Met		
800	805	810
Leu Ser Ser Gln Ala Leu Arg Leu Asp Glu Thr Gln Glu Ala Glu		
815	820	825
Phe Gln Ala Leu Arg Gln Gln Leu Gln Gln Glu Leu Glu Leu Leu		
830	835	840
Asn Ala Tyr Gln Ser Lys Ile Lys Ile Arg Thr Glu Ser Gln His		
845	850	855
Glu Arg Glu Leu Arg Glu Leu Glu Gln Arg Val Ala Leu Arg Arg		
860	865	870
Ala Leu Leu Glu Gln Arg Val Glu Glu Glu Leu Leu Ala Leu Gln		
875	880	885
Thr Gly Arg Ser Glu Arg Ile Arg Ser Leu Leu Glu Arg Gln Ala		
890	895	900
Arg Glu Ile Glu Ala Phe Asp Ala Glu Ser Met Arg Leu Gly Phe		
905	910	915
Ser Ser Met Ala Leu Gly Gly Ile Pro Ala Glu Ala Ala Ala Gln		
920	925	930
Gly Tyr Pro Ala Pro Pro Pro Ala Pro Ala Trp Pro Ser Arg Pro		
935	940	945
Val Pro Arg Ser Gly Ala His Trp Ser His Gly Pro Pro Pro Pro		
950	955	960
Gly Met Pro Pro Pro Ala Trp Arg Gln Pro Ser Leu Leu Ala Pro		
965	970	975
Pro Gly Pro Pro Asn Trp Leu Gly Pro Pro Thr Gln Ser Gly Thr		
980	985	990
Pro Arg Gly Gly Ala Leu Leu Leu Leu Arg Asn Ser Pro Gln Pro		
995	1000	1005
Leu Arg Arg Ala Ala Ser Gly Gly Ser Gly Ser Glu Asn Val Gly		
1010	1015	1020
Pro Pro Ala Ala Ala Val Pro Gly Pro Leu Ser Arg Ser Thr Ser		
1025	1030	1035
Val Ala Ser His Ile Leu Asn Gly Ser Ser His Phe Tyr Ser		
1040	1045	

<210> 8  
<211> 322  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3339024CD1

<400> 8  
Met Pro Thr Phe Ser Ile Pro Gly Thr Leu Glu Ser Gly His Pro  
1 5 10 15  
Arg Asn Leu Thr Cys Ser Val Pro Trp Ala Cys Glu Gln Gly Thr  
20 25 30  
Pro Pro Thr Ile Thr Trp Met Gly Ala Ser Val Ser Ser Leu Asp  
35 40 45  
Pro Thr Ile Thr Arg Ser Ser Met Leu Ser Leu Ile Pro Gln Pro  
50 55 60  
Gln Asp His Gly Thr Ser Leu Thr Cys Gln Val Thr Leu Pro Gly

65	70	75												
Ala	Gly	Val	Thr	Met	Thr	Arg	Ala	Val	Arg	Leu	Asn	Ile	Ser	Tyr
80									85					90
Pro	Pro	Gln	Asn	Leu	Thr	Met	Thr	Val	Phe	Gln	Gly	Asp	Gly	Thr
95									100					105
Ala	Ser	Thr	Thr	Leu	Arg	Asn	Gly	Ser	Ala	Leu	Ser	Val	Leu	Glu
110									115					120
Gly	Gln	Ser	Leu	His	Leu	Val	Cys	Ala	Val	Asp	Ser	Asn	Pro	Pro
125									130					135
Ala	Arg	Leu	Ser	Trp	Thr	Trp	Gly	Ser	Leu	Thr	Leu	Ser	Pro	Ser
140									145					150
Gln	Ser	Ser	Asn	Leu	Gly	Val	Leu	Glu	Leu	Pro	Arg	Val	His	Val
155									160					165
Lys	Asp	Glu	Gly	Glu	Phe	Thr	Cys	Arg	Ala	Gln	Asn	Pro	Leu	Gly
170									175					180
Ser	Gln	His	Ile	Ser	Leu	Ser	Leu	Ser	Leu	Gln	Asn	Glu	Tyr	Thr
185									190					195
Gly	Lys	Met	Arg	Pro	Ile	Ser	Gly	Val	Thr	Leu	Gly	Ala	Phe	Gly
200									205					210
Gly	Ala	Gly	Ala	Thr	Ala	Leu	Val	Phe	Leu	Tyr	Phe	Cys	Ile	Ile
215									220					225
Phe	Val	Val	Val	Arg	Ser	Cys	Arg	Lys	Lys	Ser	Ala	Arg	Pro	Ala
230									235					240
Val	Gly	Val	Gly	Asp	Thr	Gly	Met	Glu	Asp	Ala	Asn	Ala	Val	Trp
245									250					255
Gly	Ser	Ala	Ser	Gln	Gly	Pro	Leu	Ile	Glu	Ser	Pro	Ala	Asp	Asp
260									265					270
Ser	Pro	Pro	His	His	Ala	Pro	Pro	Ala	Leu	Ala	Thr	Pro	Ser	Pro
275									280					285
Glu	Glu	Gly	Glu	Ile	Gln	Tyr	Ala	Ser	Leu	Ser	Phe	His	Lys	Ala
290									295					300
Arg	Pro	Gln	Tyr	Pro	Gln	Glu	Gln	Glu	Ala	Ile	Gly	Tyr	Glu	Tyr
305									310					315
Ser	Glu	Ile	Asn	Ile	Pro	Lys								
320														

<210> 9  
 <211> 1212  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 4436929CD1

<400> 9														
Met	Ala	Asn	Asp	Ser	Pro	Ala	Lys	Ser	Leu	Val	Asp	Ile	Asp	Leu
1					5				10					15
Ser	Ser	Leu	Arg	Asp	Pro	Ala	Gly	Ile	Phe	Glu	Leu	Val	Glu	Val
					20				25					30
Val	Gly	Asn	Gly	Thr	Tyr	Gly	Gln	Val	Tyr	Lys	Gly	Arg	His	Val
					35				40					45
Lys	Thr	Gly	Gln	Leu	Ala	Ala	Ile	Lys	Val	Met	Asp	Val	Thr	Glu
					50				55					60
Asp	Glu	Glu	Glu	Ile	Lys	Leu	Glu	Ile	Asn	Met	Leu	Lys	Lys	
					65				70					75
Tyr	Ser	His	His	Arg	Asn	Ile	Ala	Thr	Tyr	Tyr	Gly	Ala	Phe	Ile
					80				85					90
Lys	Lys	Ser	Pro	Pro	Gly	His	Asp	Asp	Gln	Leu	Trp	Leu	Val	Met
					95				100					105
Glu	Phe	Cys	Gly	Ala	Gly	Ser	Ile	Thr	Asp	Leu	Val	Lys	Asn	Thr
					110				115					120
Lys	Gly	Asn	Thr	Leu	Lys	Glu	Asp	Trp	Ile	Ala	Tyr	Ile	Ser	Arg
					125				130					135
Glu	Ile	Leu	Arg	Gly	Leu	Ala	His	Leu	His	Ile	His	His	Val	Ile
					140				145					150
His	Arg	Asp	Ile	Lys	Gly	Gln	Asn	Val	Leu	Leu	Thr	Glu	Asn	Ala

	155		160		165									
Glu	Val	Lys	Leu	Val	Asp	Phe	Gly	Val	Ser	Ala	Gln	Leu	Asp	Arg
	170					175								180
Thr	Val	Gly	Arg	Arg	Asn	Thr	Phe	Ile	Gly	Thr	Pro	Tyr	Trp	Met
	185					190								195
Ala	Pro	Glu	Val	Ile	Ala	Cys	Asp	Glu	Asn	Pro	Asp	Ala	Thr	Tyr
	200								205					210
Asp	Tyr	Arg	Ser	Asp	Leu	Trp	Ser	Cys	Gly	Ile	Thr	Ala	Ile	Glu
	215					220								225
Met	Ala	Glu	Gly	Ala	Pro	Pro	Leu	Cys	Asp	Met	His	Pro	Met	Arg
	230					235								240
Ala	Leu	Phe	Leu	Ile	Pro	Arg	Asn	Pro	Pro	Pro	Arg	Leu	Lys	Ser
	245					250								255
Lys	Lys	Trp	Ser	Lys	Lys	Phe	Phe	Ser	Phe	Ile	Glu	Gly	Cys	Leu
	260					265								270
Val	Lys	Asn	Tyr	Met	Gln	Arg	Pro	Ser	Thr	Glu	Gln	Leu	Leu	Lys
	275					280								285
His	Pro	Phe	Ile	Arg	Asp	Gln	Pro	Asn	Glu	Arg	Gln	Val	Arg	Ile
	290					295								300
Gln	Leu	Lys	Asp	His	Ile	Asp	Arg	Thr	Arg	Lys	Lys	Arg	Gly	Glu
	305					310								315
Lys	Asp	Glu	Thr	Glu	Tyr	Glu	Tyr	Ser	Gly	Ser	Glu	Glu	Glu	
	320					325								330
Glu	Glu	Val	Pro	Glu	Gln	Glu	Gly	Glu	Pro	Ser	Ser	Ile	Val	Asn
	335					340								345
Val	Pro	Gly	Glu	Ser	Thr	Leu	Arg	Arg	Asp	Phe	Leu	Arg	Leu	Gln
	350					355								360
Gln	Glu	Asn	Lys	Glu	Arg	Ser	Glu	Ala	Leu	Arg	Arg	Gln	Gln	Leu
	365					370								375
Leu	Gln	Glu	Gln	Gln	Leu	Arg	Glu	Gln	Glu	Glu	Tyr	Lys	Arg	Gln
	380					385								390
Leu	Leu	Ala	Glu	Arg	Gln	Lys	Arg	Ile	Glu	Gln	Gln	Lys	Glu	Gln
	395					400								405
Arg	Arg	Arg	Leu	Glu	Glu	Gln	Gln	Arg	Arg	Glu	Arg	Glu	Ala	Arg
	410					415								420
Arg	Gln	Gln	Glu	Arg	Glu	Gln	Arg	Arg	Arg	Glu	Gln	Glu	Glu	Lys
	425					430								435
Arg	Arg	Leu	Glu	Glu	Leu	Glu	Arg	Arg	Arg	Lys	Glu	Glu	Glu	
	440					445								450
Arg	Arg	Arg	Ala	Glu	Glu	Glu	Lys	Arg	Val	Glu	Arg	Glu	Gln	
	455					460								465
Glu	Tyr	Ile	Arg	Arg	Gln	Leu	Glu	Glu	Glu	Gln	Arg	His	Leu	Glu
	470					475								480
Val	Leu	Gln	Gln	Gln	Leu	Leu	Gln	Glu	Gln	Ala	Met	Leu	Leu	His
	485					490								495
Asp	His	Arg	Arg	Pro	His	Pro	Gln	His	Ser	Gln	Gln	Pro	Pro	Pro
	500					505								510
Pro	Gln	Gln	Glu	Arg	Ser	Lys	Pro	Ser	Phe	His	Ala	Pro	Glu	Pro
	515					520								525
Lys	Ala	His	Tyr	Glu	Pro	Ala	Asp	Arg	Ala	Arg	Glu	Val	Glu	Asp
	530					535								540
Arg	Phe	Arg	Lys	Thr	Asn	His	Ser	Ser	Pro	Glu	Ala	Gln	Ser	Lys
	545					550								555
Gln	Thr	Gly	Arg	Val	Leu	Glu	Pro	Pro	Val	Pro	Ser	Arg	Ser	Glu
	560					565								570
Ser	Phe	Ser	Asn	Gly	Asn	Ser	Glu	Ser	Val	His	Pro	Ala	Leu	Gln
	575					580								585
Arg	Pro	Ala	Glu	Pro	Gln	Val	Pro	Val	Arg	Thr	Thr	Ser	Arg	Ser
	590					595								600
Pro	Val	Leu	Ser	Arg	Arg	Asp	Ser	Pro	Leu	Gln	Gly	Ser	Gly	Gln
	605					610								615
Gln	Asn	Ser	Gln	Ala	Gly	Gln	Arg	Asn	Ser	Thr	Ser	Ser	Ile	Glu
	620					625								630
Pro	Arg	Leu	Leu	Trp	Glu	Arg	Val	Glu	Lys	Leu	Val	Pro	Arg	Pro
	635					640								645
Gly	Ser	Gly	Ser	Ser	Ser	Gly	Ser	Ser	Asn	Ser	Gly	Ser	Gln	Pro
	650					655								660

Gly Ser His Pro Gly Ser Gln Ser Gly Ser Gly Glu Arg Phe Arg  
 665 670 675  
 Val Arg Ser Ser Ser Lys Ser Glu Gly Ser Pro Ser Gln Arg Leu  
 680 685 690  
 Glu Asn Ala Val Lys Lys Pro Glu Asp Lys Lys Glu Val Phe Arg  
 695 700 705  
 Pro Leu Lys Pro Ala Gly Glu Val Asp Leu Thr Ala Leu Ala Lys  
 710 715 720  
 Glu Leu Arg Ala Val Glu Asp Val Arg Pro Pro His Lys Val Thr  
 725 730 735  
 Asp Tyr Ser Ser Ser Ser Glu Glu Ser Gly Thr Thr Asp Glu Glu  
 740 745 750  
 Asp Asp Asp Val Glu Gln Glu Gly Ala Asp Glu Ser Thr Ser Gly  
 755 760 765  
 Pro Glu Asp Thr Arg Ala Ala Ser Ser Leu Asn Leu Ser Asn Gly  
 770 775 780  
 Glu Thr Glu Ser Val Lys Thr Met Ile Val His Asp Asp Val Glu  
 785 790 795  
 Ser Glu Pro Ala Met Thr Pro Ser Lys Glu Gly Thr Leu Ile Val  
 800 805 810  
 Arg Gln Thr Gln Ser Ala Ser Ser Thr Leu Gln Lys His Lys Ser  
 815 820 825  
 Ser Ser Ser Phe Thr Pro Phe Ile Asp Pro Arg Leu Leu Gln Ile  
 830 835 840  
 Ser Pro Ser Ser Gly Thr Thr Val Thr Ser Val Val Gly Phe Ser  
 845 850 855  
 Cys Asp Gly Met Arg Pro Glu Ala Ile Arg Gln Asp Pro Thr Arg  
 860 865 870  
 Lys Gly Ser Val Val Asn Val Asn Pro Thr Asn Thr Arg Pro Gln  
 875 880 885  
 Ser Asp Thr Pro Glu Ile Arg Lys Tyr Lys Lys Arg Phe Asn Ser  
 890 895 900  
 Glu Ile Leu Cys Ala Ala Leu Trp Gly Val Asn Leu Leu Val Gly  
 905 910 915  
 Thr Glu Ser Gly Leu Met Leu Leu Asp Arg Ser Gly Gln Gly Lys  
 920 925 930  
 Val Tyr Pro Leu Ile Asn Arg Arg Arg Phe Gln Gln Met Asp Val  
 935 940 945  
 Leu Glu Gly Leu Asn Val Leu Val Thr Ile Ser Gly Lys Lys Asp  
 950 955 960  
 Lys Leu Arg Val Tyr Tyr Leu Ser Trp Leu Arg Asn Lys Ile Leu  
 965 970 975  
 His Asn Asp Pro Glu Val Glu Lys Lys Gln Gly Trp Thr Thr Val  
 980 985 990  
 Gly Asp Leu Glu Gly Cys Val His Tyr Lys Val Val Lys Tyr Glu  
 995 1000 1005  
 Arg Ile Lys Phe Leu Val Ile Ala Leu Lys Ser Ser Val Glu Val  
 1010 1015 1020  
 Tyr Ala Trp Ala Pro Lys Pro Tyr His Lys Phe Met Ala Phe Lys  
 1025 1030 1035  
 Ser Phe Gly Glu Leu Val His Lys Pro Leu Leu Val Asp Leu Thr  
 1040 1045 1050  
 Val Glu Glu Gly Gln Arg Leu Lys Val Ile Tyr Gly Ser Cys Ala  
 1055 1060 1065  
 Gly Phe His Ala Val Asp Val Asp Ser Gly Ser Val Tyr Asp Ile  
 1070 1075 1080  
 Tyr Leu Pro Thr His Ile Gln Cys Ser Ile Lys Pro His Ala Ile  
 1085 1090 1095  
 Ile Ile Leu Pro Asn Thr Asp Gly Met Glu Leu Leu Val Cys Tyr  
 1100 1105 1110  
 Glu Asp Glu Gly Val Tyr Val Asn Thr Tyr Gly Arg Ile Thr Lys  
 1115 1120 1125  
 Asp Val Val Leu Gln Trp Gly Glu Met Pro Thr Ser Val Ala Tyr  
 1130 1135 1140  
 Ile Arg Ser Asn Gln Thr Met Gly Trp Gly Glu Lys Ala Ile Glu  
 1145 1150 1155  
 Ile Arg Ser Val Glu Thr Gly His Leu Asp Gly Val Phe Met His

1160 Lys Arg Ala Gln Arg Leu Lys Phe Leu Cys 1175 Val Phe Phe Ala Ser Val Arg Ser Gly Gly 1190 Phe Met Thr Leu Gly Arg Thr Ser Leu Leu 1205	1165 Glu Arg Asn Asp Lys 1180 Ser Ser Gln Val Tyr 1195 Ser Trp 1210	1170 1185 1200
---	---	----------------------

<210> 10  
<211> 280  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 5046791CD1

<400> 10 Met Gln Pro Leu Arg Val Asn Ser Gln Pro Gly Pro Gln Lys Arg 1 5 10 15 Cys Leu Phe Val Cys Arg His Gly Glu Arg Met Asp Val Val Phe 20 25 30 Gly Lys Tyr Trp Leu Ser Gln Cys Phe Asp Ala Lys Gly Arg Tyr 35 40 45 Ile Arg Thr Asn Leu Asn Met Pro His Ser Leu Pro Gln Arg Ser 50 55 60 Gly Gly Phe Arg Asp Tyr Glu Lys Asp Ala Pro Ile Thr Val Phe 65 70 75 Gly Cys Met Gln Ala Arg Leu Val Gly Glu Ala Leu Leu Glu Ser 80 85 90 Asn Thr Ile Ile Asp His Val Tyr Cys Ser Pro Ser Leu Arg Cys 95 100 105 Val Gln Thr Ala His Asn Ile Leu Lys Gly Leu Gln Gln Glu Asn 110 115 120 His Leu Lys Ile Arg Val Glu Pro Gly Leu Phe Glu Trp Thr Lys 125 130 135 Trp Val Ala Gly Ser Thr Leu Pro Ala Trp Ile Pro Pro Ser Glu 140 145 150 Leu Ala Ala Ala Asn Leu Ser Val Asp Thr Thr Tyr Arg Pro His 155 160 165 Ile Pro Ile Ser Lys Leu Val Val Ser Glu Ser Tyr Asp Thr Tyr 170 175 180 Ile Ser Arg Ser Phe Gln Val Thr Lys Glu Ile Ile Ser Glu Cys 185 190 195 Lys Ser Lys Gly Asn Asn Ile Leu Ile Val Ala His Ala Ser Ser 200 205 210 Leu Glu Ala Cys Thr Cys Gln Leu Gln Gly Leu Ser Pro Gln Asn 215 220 225 Ser Lys Asp Phe Val Gln Met Val Arg Lys Ile Pro Tyr Leu Gly 230 235 240 Phe Cys Ser Cys Glu Glu Leu Gly Glu Thr Gly Ile Trp Gln Leu 245 250 255 Thr Asp Pro Pro Ile Leu Pro Leu Thr His Gly Pro Thr Gly Gly 260 265 270 Phe Asn Trp Arg Glu Thr Leu Leu Gln Glu 275 280
--

<210> 11  
<211> 114  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 1416174CD1

<400> 11

Met	Leu	Ala	Ile	Ser	Pro	Ser	His	Leu	Gly	Ala	Asp	Leu	Val	Ala
1								5	10					15
Ala	Pro	His	Ala	Arg	Phe	Asp	Asp	Gly	Leu	Val	His	Leu	Cys	Trp
								20	25					30
Val	Arg	Thr	Gly	Ile	Ser	Arg	Ala	Ala	Leu	Leu	Arg	Leu	Phe	Leu
								35	40					45
Ala	Met	Glu	Arg	Gly	Ser	His	Phe	Ser	Leu	Gly	Cys	Pro	Gln	Leu
								50	55					60
Gly	Tyr	Ala	Ala	Ala	Arg	Ala	Phe	Arg	Leu	Glu	Pro	Leu	Thr	Pro
								65	70					75
Arg	Gly	Val	Leu	Thr	Val	Asp	Gly	Glu	Gln	Val	Glu	Tyr	Gly	Pro
								80	85					90
Leu	Gln	Ala	Gln	Met	His	Pro	Gly	Ile	Gly	Thr	Leu	Leu	Thr	Gly
								95	100					105
Pro	Pro	Gly	Cys	Pro	Gly	Arg	Glu	Pro						
								110						

<210> 12  
<211> 375  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3244919CD1

<400> 12															
Met	Gly	Ser	Ser	Met	Ser	Ala	Ala	Thr	Ala	Arg	Arg	Pro	Val	Phe	
1									5	10				15	
Asp	Asp	Lys	Glu	Asp	Val	Asn	Phe	Asp	His	Phe	Gln	Ile	Leu	Arg	
								20	25					30	
Ala	Ile	Gly	Lys	Gly	Ser	Phe	Gly	Lys	Val	Cys	Ile	Val	Gln	Lys	
								35	40					45	
Arg	Asp	Thr	Glu	Lys	Met	Tyr	Ala	Met	Lys	Tyr	Met	Asn	Lys	Gln	
								50	55					60	
Gln	Cys	Ile	Glu	Arg	Asp	Glu	Val	Arg	Asn	Val	Phe	Arg	Glu	Leu	
								65	70					75	
Glu	Ile	Leu	Gln	Glu	Ile	Glu	His	Val	Phe	Leu	Val	Asn	Leu	Trp	
								80	85					90	
Tyr	Ser	Phe	Gln	Asp	Glu	Glu	Asp	Met	Phe	Met	Val	Val	Asp	Leu	
								95	100					105	
Leu	Leu	Gly	Gly	Asp	Leu	Arg	Tyr	His	Leu	Gln	Gln	Asn	Val	Gln	
								110	115					120	
Phe	Ser	Glu	Asp	Thr	Val	Arg	Leu	Tyr	Ile	Cys	Glu	Met	Ala	Leu	
								125	130					135	
Ala	Leu	Asp	Tyr	Leu	Arg	Gly	Gln	His	Ile	Ile	His	Arg	Asp	Val	
								140	145					150	
Lys	Pro	Asp	Asn	Ile	Leu	Leu	Asp	Glu	Arg	Gly	His	Ala	His	Leu	
								155	160					165	
Thr	Asp	Phe	Asn	Ile	Ala	Thr	Ile	Ile	Lys	Asp	Gly	Glu	Arg	Ala	
								170	175					180	
Thr	Ala	Leu	Ala	Gly	Thr	Lys	Pro	Tyr	Met	Ala	Pro	Glu	Ile	Phe	
								185	190					195	
His	Ser	Phe	Val	Asn	Gly	Gly	Thr	Gly	Tyr	Ser	Phe	Glu	Val	Asp	
								200	205					210	
Trp	Trp	Ser	Val	Gly	Val	Met	Ala	Tyr	Glu	Leu	Leu	Arg	Gly	Trp	
								215	220					225	
Arg	Pro	Tyr	Asp	Ile	His	Ser	Ser	Asn	Ala	Val	Glu	Ser	Leu	Val	
								230	235					240	
Gln	Leu	Phe	Ser	Thr	Val	Ser	Val	Gln	Tyr	Val	Pro	Thr	Trp	Ser	
								245	250					255	
Lys	Glu	Met	Val	Ala	Leu	Leu	Arg	Lys	Leu	Leu	Thr	Val	Asn	Pro	
								260	265					270	
Glu	His	Arg	Leu	Ser	Ser	Leu	Gln	Asp	Val	Gln	Ala	Ala	Pro	Ala	
								275	280					285	
Leu	Ala	Gly	Val	Leu	Trp	Asp	His	Leu	Ser	Glu	Lys	Arg	Val	Glu	
								290	295					300	

Pro	Gly	Phe	Val	Pro	Asn	Lys	Gly	Arg	Leu	His	Cys	Asp	Pro	Thr
				305					310					315
Phe	Glu	Leu	Glu	Glu	Met	Ile	Leu	Glu	Ser	Arg	Pro	Leu	His	Lys
				320					325					330
Lys	Lys	Lys	Arg	Leu	Ala	Lys	Asn	Lys	Ser	Arg	Asp	Asn	Ser	Arg
				335					340					345
Asp	Ser	Ser	Gln	Ser	Ala	Pro	Arg	Ser	Lys	Ser	Lys	Pro	Ser	Thr
				350					355					360
Gln	Arg	Gln	Gly	Ser	Trp	Ala	Leu	Ala	Ser	Ser	Gly	Leu	Gly	Glut
				365					370					375

<210> 13  
<211> 1859  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 058860CB1

<210> 14  
<211> 3501  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 2041716CB1

<400> 14  
gtgggtggc tgcaagtggag agttcccaac aaggctacgc agaagaaccc cttgactga 60  
agcaatggag ggggggtccag ctgtctgctg ccaggatcct cgggcagagc tggttagaacg 120  
qgtggcagcc atcgatgtqa ctcacttqqa qgagggcagat qgtggcccaq aqccatacq 180

aaacggtgtg	gaccccccac	cacggggcag	agctgcctct	gtgatccctg	gcagtacttc	240
aagactgctc	ccagcccgac	ctagcctctc	agccaggaag	ctttccctac	aggagcggcc	300
agcaggaagc	tatctggagg	cgcaggtctgg	gccttatgcc	acggggcctg	ccagccacat	360
ctcccccccg	gcctggcgga	ggcccaccat	cgagtcccac	cacgtggcca	tctcagatgc	420
agaggactgc	gtcagactga	accagtacaa	gctgcagagt	gagatggca	aggtggggct	480
gactgatgcc	tatctgcagg	gtgcctacgg	tgtggtgagg	ctggcctaca	acgaaaagtga	540
agacagacac	tatgcaatga	aagtcccttc	caaaaagaag	ttactgaagc	agtatggctt	600
tccacgtcgc	cctcccccg	gagggttcca	ggctgcccag	ggaggaccag	ccaagcagct	660
gctgcccctg	gagcgggtgt	accaggagat	tgccatcctg	aagaagctgg	accacgtgaa	720
tgtggtcaaa	ctgatcgagg	tcctggatga	cccagctgag	gacaacctct	atttggttga	780
cctcctgaga	aaggggcccg	tcatggaaat	gcccgtgtac	aaggccttct	cggaggagca	840
agctcgccctc	tacctgcggg	acgtcatcct	ggcctcgag	tacttgcact	gccagaagat	900
cgtccacacgg	gacatcaagc	catccaaac	gtcctgggg	gatgatgggc	acgtgaagat	960
cgccgacttgc	ggcgtcagca	accagttga	ggggaacgcac	gctcagctgt	ccagcacggc	1020
gggaacccca	gcattcatgg	ccccggggc	catttctgtat	tccggccaga	gcttcagttg	1080
gaaggccttg	gatgtatggg	ccactggcgt	cacgttgcac	tgctttgtct	atgggaagtg	1140
cccggttcatc	gacgatttca	tcctggccct	ccacaggaag	atcaagaatg	agcccgttgt	1200
gtttccttag	gagccagaaa	tcagcgagga	gotcaaggac	ctgatcctga	agatgttaga	1260
caagaatccc	gagacgagaa	ttggggtgcc	agacatcaag	ttgcaccctt	gggtgaccaa	1320
gaacggggag	gagccccttc	cttcggagga	ggagcactgc	agcgtgggtt	aggtgacaga	1380
ggaggaggtt	aagaactcag	tcaggctcat	ccccagctgg	accacgggtga	tcctgggtaa	1440
gtccatgtgc	aggaagcggt	ccttggaa	cccggtttag	ccccaaagcac	ggagggaaaga	1500
gcatccatgc	tctgtctccag	gaaacactt	ggtgaaagaa	gggtttgggt	aagggggcaaa	1560
gagcccgagag	tcccccggcg	tccaggaaga	cgaggctgca	tcctgagcccc	ctgcatgcac	1620
ccaggggccac	ccggcagcac	actatcccg	cggctccaga	ggcccaccccc	tcatgcaaca	1680
gccgcccccc	caggcagggg	gctggggact	gcagccccac	tcccgccccct	cccccatcg	1740
gctgcatgac	ctccacgcac	gcacgtccag	ggacagactg	aatgttatgt	catttgggt	1800
cttggggggca	gggctccac	gaggccatcc	tcctcttctt	ggacccctt	ggcctgaccc	1860
attctgtggg	gaaaccgggt	gcccattggag	cctcagaaat	gccacccggc	tgggtggcat	1920
ggcctggggc	aggaggcaga	ggcaggagac	caagatggca	ggtggaggcc	aggcttacca	1980
caacggaaaga	gaccccccgc	tggggccggg	caggcctggc	tcagctgcca	caggcatatg	2040
gtggagaggg	gggttacccgt	cccacccctgg	ggtgggtggca	ccagagctct	tgtcttattca	2100
gacgctggta	tggggctcg	gacccttac	tggggcagg	gccagtgtt	gagaattctg	2160
attctttttt	tgtttctttt	tactttttt	ttaaacctgg	gggttctgggg	agaggccctg	2220
cttgggaaca	tctcacgagc	tttccttacat	cttcctgtt	tcccagcaca	gccccaaaggatt	2280
atttggcagc	caagtggatg	gaactaactt	tcctggactg	tgtttcgcat	tcggcggttat	2340
ctggaaagtgc	gactgaacgg	aatcaagctc	tgagcagagg	cctgaagcgg	aagcaccacaca	2400
tcgtccctgc	ccatctcaact	ctctcccttgc	atgatcccc	tagagctgag	gctggagaag	2460
acaccaggggc	tgacttttgc	cgagggccat	ggacgcgaca	ggcctgtggc	cctgcgcgt	2520
ctgaaataac	tggaaacccag	cctctcttcc	tacaccggcc	taccatctg	ggcccaaaagag	2580
ctgcactcac	actcttacaa	cgaaggacaa	actgtccagg	tcggaggat	cacgagacac	2640
agaacacttgg	gggggtgtca	cgctggcagg	tggcctctgc	ggcaatttgc	tcacccttag	2700
gacatcagca	gtcagctgc	tcaaggggg	ggtgtccgg	cgcgtcgaga	cacagcttt	2760
ccggagcagc	cttcacccctc	tctctggat	cagtgtccgg	ctggccgacg	tggcatttgc	2820
tgaccgaatg	ctcatagagg	ttgaccccca	cagggtcacg	caggactcgg	acactgcct	2880
gaaacatgg	atggacaagg	gcttttgcc	acaggtgtgg	gtgtctgtt	ggaggaggggc	2940
ttgtttggag	aaggagggtt	ggctggggga	gaaaccggc	tcccgtgca	tctccgcgccc	3000
tgtgggtgca	tgtcgcgtgc	tcatctgttgc	cacacagctc	actcttatgt	cctgcactgg	3060
tacatgcac	tgtatatacg	tttctacgtc	tatataaggc	taggagccga	atgtgcccc	3120
ttgtcagttg	gtccacgttt	ctccccccgt	cctctgggtt	aaggcagtgt	ggcccgaggc	3180
ttaaaaagg	actctgttttact	gttttaaga	acacttttat	agaggtagtg	gaaggcaagt	3240
taagagccaa	tcaactgtatcc	ccaagtgttt	ctttagacatc	tgggtctgggg	ggaccacatt	3300
gatcgaccc	acccttggaa	agctcagggg	taggcccagg	ttggatgtctc	accctgtcac	3360
tgagggtttt	ggttggcattc	gttggtttttgc	aatgttagcac	aagcgatgag	caaactctat	3420
aagagtgttt	taaaaaattaa	cttcccagga	agttagttaa	aaacaataaa	agcccttct	3480
tgagttaaaa	agaaaaaaa	a				3501

<210> 15  
 <211> 3039  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7472005CB1

<400> 15

atggcccccg cccggggccg cctgccccct gcgcctctggg tcgtcacggc cgccggccgc 60  
gcggccacct gcgtgtccgc ggccgcgcgc gaagtgaatt tgctggacac gtcgaccatc 120  
cacggggact ggggctggct cacgtatccg gtcatgggt gggactccat caacgagggt 180  
gacgactct tccagcccat ccacacgtac caggtttgca acgtcatgag ccccaaccag 240  
aacaactggc tgcgcacgag ctgggtcccc cgagacggcg cccggcgcgt ctatgctgag 300  
atcaagttt ccctgcgcga ctgcaacagc atgcctgggt tgctgggcac ctgcaaggag 360  
accttcaacc tctactacct ggagtcggac cgcgcacctgg gggcagcac acaagaaaagc 420  
cagttcctca aaatcgacac cattgcggcc gacgagagct tcacaggtgc cgaccttgg 480  
gtgcggcgtc tcaagctcaa cacggaggtg cgcaagtgtgg gtccctctcg caagcgcggc 540  
ttctacctgg ccttccagga cataggtgcc tgcctggcca tcctctctc ccgcacatc 600  
tataagaagt gcccgtccat ggtgcgaat ctggctgcct tctcgagggc agtgcacgggg 660  
gccgactcgt cctcactggc gacacacca agatgtactg gtgtgcgtgg cccgctacga 720  
ggaggtgagg gcccagtgcg tgccggactc agaggagcgg 780  
cagcgcggg ggcgagtggc tcgtgcccatt gggccatgc 840  
tacaagtcag cccctgggg gcaactgtgtg cccgcgtgc ccaactccgc 900  
gctccagccg cccaaaggctg ccaactgtac accgtgcagc cctggaccgg 960  
ccgtcctca gctgcacccg gggacatca g tgcactctgg acctacaatg ccgtgtgcc 1020  
acgggcaccc gctttgtgcc aacctgctgg cccacatgaa ctactcctc tgatcgagg cgtgtccgac 1080  
ctgagccccc agcccccggc tcccagggtgg tggtgatccg gggacatca gcaactgtac 1140  
caggagcccg agcagccgaa gacaaggaga tgcagagcta gcaactgtac accgtgcac 1200  
ggcctcaagc cgggcacccg ggccgcttca gccaggccat aggaccattg tctggatctg 1260  
ctcatctgca agaagaggca aagatgcaatc atcagaatgg ctcacccctc tgatcgagg 1320  
ccggggaaagc tccagagcc cggggggggc gcagttcac atcggctctg gagactccgg 1380  
atcggctctg tggccatcaa tcaagagccgg gggccatca gcaactgtac accgtgcac 1440  
gtatgtggccg tggccatcaa tccacccctc aaggccgtca ccaccaagac caccgtctcc 1500  
tgcgttcc tcaactgttcc cagggtccgag cccgcacccg ggggggggg 1560  
ggaggtggag accggggaaac cccggccccc ctgcgttcc tcaactgttcc 1620  
cctgacgctc atcagggcc tgggtgtgt tctgctctg 1680  
ctgtggctac agcaaggcc tccaggactc ggacgaggag 1740  
acaggcaccc ccacctgtct tcctgcctct gcatcacccc 1800  
ccagttctat gggaaacccc acacccatca gggccatcc 1860  
tcgggagatc gggccctcta ggatccatcat cggccatcc 1920  
ggaagtctgc tacggggggc tgcgggtgcc agggcagcgg 1980  
ggccctaaa gccggctaca cggagagaca gggccgggac 2040  
catggggcaa ttgcaccatc ccaacatcat ccgcctcgag 2100  
cctggcaatg attgtactg agtacatgg gaaacggct 2160  
ccacgacggg cagttcacca tcacgcgt ggtggcatg 2220  
catgcgtac ctctcagacc tggctatgt ccaccgagac 2280  
gttgacagc aacctgttcc gcaagggtgc tgacttcggg 2340  
cgacccggat gtcgttccata ccaccacggg cgggaagatc 2400  
agagggccatc gcctccgc ccttctctc ggccagcgc 2460  
catgtgggg gttgtggccct atggggagcg gcccactgg 2520  
gagtgcctaa cccctggcagg tcatacgat tggggaggag 2580  
catgggtgc ccccacggcc tgcaccagct catgctcgac 2640  
gtcggcgtt cgttctccc agattgttag tgcctcgat 2700  
gagttctcagg gccaccggca cagtcacgc gttggccccc 2760  
ctttgaccc tcggggggca ggggtggcgg tgggggctc 2820  
accgtggggg actggctgga ctccatccgc atggggccgt accgagacca ttgcgttcc 2880  
ggcggatact cctctctgg catgggtcta cgcatacgc cccaggacgt ggcgcctctg 2940  
gcatcaccc tcatggggca ccagaagaag atcctggggca gcatcagac catgcgggccc 3000  
cagctgacca gcacccagggg gccccccggg cacctctga 3039

<210> 16  
<211> 1104  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7472006CB1

<400> 16  
atggatgacg ctgtgtccct caagcgtacga ggctacccctc tggggataaa tttaggagag 60  
ggctcctatg caaaaatggaa atctgtttac tctgagcgcgc tgaagttcaa tggggatgtc 120  
aagatcatcg accgcaagaa ggccccccca gacttcttgg agaaatttct tccccggaa 180  
attgagatcc tggccatgtt aaaccactgc tccatcatta agacctacga gatcttttag 240  
acatcacatg gcaagggtcta catgtcatg gagctcgcag tccaggcga ctcctcgag 300

ttaatcaaaa	cccggggagc	cctgcatgag	gacgaagctc	gcaagaagtt	ccaccagctt	360
tccttgccca	tcaagtactg	ccacgacactg	gacgtcgcc	accgggacct	caagtgtgac	420
aaccttctcc	ttgacaagga	cttcaacatc	aagctgtccg	acttcagctt	ctccaagcgc	480
tgcctgcggg	atgacagtgg	tcgaatggcc	ttaagcaaga	ccttctgtgg	gtcaccagcg	540
tatgcggccc	caaggggtgt	gcagggcatt	cccttaccagc	ccaaggtgta	cgacatctgg	600
agcctaggcg	tgatccctta	catcatggtc	tgcggctcca	tgccttacga	cgactccaac	660
atcaagaaga	tgtctgcgtat	ccagaaggag	caccgcgtca	acttcccacg	ctccaaggcac	720
ctgacaggcg	agtgcagaagg	cctcatctac	cacatgtcgc	agccccgacgt	caaccggcgg	780
ctccacatcg	agagatccct	cagccactgc	tggatgcagc	ccaaggcacg	gggatctccc	840
tctgtggcca	tcaacaagga	ggggggagatg	tcccggggaa	ctgaaccctt	gtggaccccc	900
gaacctggct	ctgacaagaa	gtctgcccacc	aagctggagc	ctgagggaga	ggcacagcccc	960
caggcacagc	ctgagacaaa	acccgagggg	acagcaatgc	aatgtccag	gcagtcggag	1020
atccctgggtt	tccccagcaa	gccgtcgact	atggagacag	aggaagggcc	cccccaacag	1080
cctccagaga	cgcggggccca	gtga				1104

<210> 17  
<211> 3939  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 2902460CB1

<400> 17  
ccgcagtgtg ctggaaaggc agctgcggca gtagcgtgag cagcccaagt tgggctggtc 60  
gcctgcgagg ggaccggcag caggtggtgg cagccgtac cctctccccg ccagggccgga 120  
ggaggccaag aggaagctgc ggatcttgc ggcgcgaggta cagaacgtgc aggtgaacca 180  
aaaagtggc atgtttgagg cgcacatcca ggcacagagc tccgcattc aagcccccg 240  
cagccccgt tgggcaggg ctcgcgc ctcgcgtgc cccttccgca gcagcagtca 300  
gccccctgg aaggctctgg ttccaggcgc cggagcgag gaacggagga caaagtccctg 360  
ggggagcaa tggccagaga cttcaggaac cgactccggg aggaaggagg gggccagcct 420  
atgctctcg caggtaaga aaggaatgcc acctcttccc gggccggctg cccctcagg 480  
atcagaggct cagggtccat cgcctttgt aaggatggag aagggtatcc ctgcctgtcc 540  
ccgctgtggc tcacccacag ctatggaaat tgacaaaagg ggctctcta ccccgaaac 600  
tcggagctgc cttagctccct cattgggct gttcggagct agcttaacga tggccacgga 660  
agtggcagcg agagttacat ccactgggcc acaccgtcca caggatcttgc cccctactga 720  
gccgtctggg agagccgtg agcttgagga cctgcagccc ccagaggccc tggtgagag 780  
gcaggggcag ttctggcgt gtgagacaag cccagcccca gaaaggggccc gggcccgca 840  
tggagaaccc ctggggaaaga tggggaaagg atatctgccc tggtgcatcc cgggtctgg 900  
ggagcctgaa gtggggcaaaa ggccagagga gacgactgtg agctgcaaa ggcagagtc 960  
ctctgtatccc ctggagctggt ccaggctgc gggcccaacttgc gctccgtag gcccggaga 1020  
ggcccaagt gggggccccc tggggggggg agggtccccca acgctgggtc tgcttggggg tggggagggc 1080  
tgtggaggcg gaaattcctt ctggcagaat gaaagatctg aaagaacccctc agtgccttcc cagccgggtt gcttggacgc gtggcacagg 1140  
ctccagggtt tggcagggca ccatggagaa ggtcaatca gaggggactt gggaaagcca gcccgggtt gcttggacgc gtggcacagg 1200  
ggagctgcta ccccaagatc aggacaagcc cataacctgt gtcatcattt cagacatggg tggggacagg gtgggtgtgc agcctggaa 1260  
gcagggaaagc ctcggggca acctggccct cacgggcttc tcttcatcc acgaagactt tggcaggatcc ctgaagaaggc actgtgagtc 1320  
caccctggac cccaaacttag ctttctgcata caaatcatgg aggaagataaa aaaacatggt ggtgtctg agtgcgttcc tggcaggag 1380  
gagaagatcc ccttggatcc agctggcagg ggtgtctg aaggcctgca ccccaaggag gatggggctt tggggggccct 1440  
tggcaggatcc ctgaagaaggc actgtgagtc ggtgtctg aaggcctgca caccctggac gatggggccct tggggggccct 1500  
ggatgtgtcg aggcccttcg tacctgcata caccctggac cccaaacttag ctttctgcata caaatcatgg aggaagataaa aaaacatggt gggggccct 1560  
ctacaaccatc atggacgacc tgctggccga gatgggaaatc aggacctacc tggaggagga gggggccct 1620  
gatgggaaatc aggacctacc tggaggagga gcggaaggac atgtaccaga agatgatcga aagggggccct 1680  
aaaagcacag cgggctgtga ccaagccacg cacggccacc ctgggggttca ggatcgagg ggtgtctg agtgcgttcc tggcaggag 1740  
ggacttcaag aagaccaaaaa cgaggggagca aggaaaaccat aacatcttgc tgcctatcc gatgcgttcc tggcaggag 1800  
cgacaagaag gaacaggcca aagtgtggat tggcaggatcc ctgaagaaggc actgtgagtc ggtgtctg aaggcctgca caccctggac 1860  
tgaggccag accctqagc atgacgttcc ggtgtctg aaggcctgca caccctggac 1920  
ccatggggat gtggtaaagg ctgcactcg ccctgtgtga gtcacgaaag gcccggaaaga ggtgtctg aaggcctgca caccctggac 1980  
cttcgcactcg ccctgtgtga gtcacgaaag gcccggaaaga ggtggacccc gaggccccca gtacatgcag tggcgggaga ccatcagctc 2040  
gatgggaaatc aatcaagaaa gaagacggca ggttcccgag gtcacgaaag gcccggaaaga ggtggacccc gaggccccca 2100  
ggacttcaag aagaccaaaaa cgaggggagca aggaaaaccat aacatcttgc tgcctatcc gatgcgttcc tggcaggag 2160  
cgacaagaag gaacaggcca aagtgtggat tggcaggatcc ctgaagaaggc actgtgagtc ggtgtctg aaggcctgca caccctggac 2220  
tgaggccag accctqagc atgacgttcc ggtgtctg aaggcctgca caccctggac 2280  
ccatggggat gtggtaaagg ctgcactcg ccctgtgtga gtcacgaaag gcccggaaaga ggtggacccc gaggccccca 2340  
gttgcacttt ggtcatttgc agtgcgttcc tggcaggag 2400  
gatgcgttcc tggcaggag 2460  
ctggcaggag qggaaaccqgg 2520

cctctcgaaa	ctcaataacc	tcgtcgacat	cctgaccgag	atgtcccagg	atgccccact	2580
cgcctgagct	gcccacgccc	tccctggccc	cgcctgggc	ctcccttcct	cctccctgtgc	2640
ttcctttctc	gttcttaact	tttccctcac	ttacacctga	ctgaccctcc	tgaactgcac	2700
tacaagacac	tttttagaaag	aggagatgag	agtttcttagt	cattttccta	acttcagggc	2760
ttggaggtgg	tgtttgcact	gctttttgt	gagagggtca	cctactagaa	gagaaatgcc	2820
cagtcttaga	ggtgggtcag	gtgttagagct	ggaggggggtc	cctggctgct	gaggggaccc	2880
taccagatga	gcccgtcctc	tgggagcccc	cttaggaagca	ccagcctgga	cctaccacct	2940
gcggaggcct	gctgccccct	ggcggccagt	gtgttagag	tgctgccaag	cacagcccta	3000
tttctgcccc	ggcctccca	ccggagagcc	cagggggccg	gccgggttcc	ttgtccctgg	3060
ctgggagcag	ggcttctgg	tagttgggc	acaaaaccat	cggggaacca	catgttact	3120
gtgagcaaag	tgtcttccga	ttagcagcct	cagggatgcc	ctgggtggct	ctccagggtct	3180
gctcaggca	ggccccccac	ccatctggta	tggaaacctg	ccggctccag	gccagaccca	3240
ggagccaaga	gaaggctgaa	gccagcttgg	ctgtgttctc	tgatcttagc	cttcccagag	3300
gaggcgagca	gaagctgtgc	cacttggat	tgcaaccat	gattcagaa	ggcacactct	3360
gccatgctga	gctccaagggg	tgctaccagg	ggaagatggg	atctatagag	tctctgggccc	3420
ctggccccag	ggaggagcac	atttttcttg	accctcacct	acctgggtct	agttggtcaa	3480
ccctgcctgc	atacatgggc	tcctgtcatg	gggcccagag	tcccttgcag	atataaaaaat	3540
agggggaggag	ctcaggctcg	cgccaggcag	gaagaaggca	ggcttctggc	ttccagaggt	3600
gccgcgggtgg	cctcctggca	tcatttgtta	ttgcctctga	aacaagcctt	actgcctgga	3660
gggcttagat	tcctgcttct	ccaatgttagt	gtgggtatct	tgttagggtat	gtgggtggatg	3720
caaggcgtg	ctccaggcac	ctcttctgta	agtctctgca	tttgagatt	cgtggagaac	3780
ctatthaagc	ccaaatttaa	ctgaaagcca	gtgagctgtga	tatgaaaggg	aatgtaaaaat	3840
ttgcctgact	tcttaagaac	aaaacccca	gtctgtgccc	ccatgctcct	tggggcttgc	3900
cacccactcc	tttgctgtca	gaggtacagg	agctgggag			3939

&lt;210&gt; 18

&lt;211&gt; 1381

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6383934CB1

&lt;400&gt; 18

atgaggacaa	tgcctgctgg	cccacatgac	ggggggatgt	agacggcagc	ggcgccagtc	60
gctcctggca	ccatggacga	tgccacagtc	ctaaggaaga	agggttacat	cgtaggcattc	120
aatcttggca	agggttccta	cgcaaaaagtc	aaatctgcct	actctgagcg	cctcaagttc	180
aatgtggctg	tcaagatcat	cgaccgcaag	aaaacaccta	ctgactttgt	ggagagattc	240
cttcctcggt	agatggacat	cctggcaact	gtcaaccacg	gctccatcat	caagacttac	300
gagatctttg	agacctctga	cggacggatc	tacatcatca	tggagcttgg	cgtccagggc	360
gacctctcg	agttcatcaa	gtgcccggga	gccctgcatg	aggacgtggc	acgcaagatg	420
ttccgacagc	tctcctccgc	cgtcaagtac	tgccacgacc	tggacatctgt	ccaccgggac	480
ctcaagtgcg	agaacattct	cctcgacaag	gacttcaaca	tcaagctgtc	tgactttggc	540
ttctccaagc	gctgcctgcg	ggacagcaat	gggcgcatca	tcctcagcaa	gaccttctgc	600
gggtcggcag	catatgcagc	ccccgaggtg	ctgcagagca	tcccctacca	gcccaggtg	660
tatgacatct	ggagcctagg	cgtgatctc	tacatcatgg	tctgcggctc	catgccctac	720
gacgactccg	acatcaagaa	gatgctgcgt	atccagaagg	agcaccgcgt	caacttccca	780
cgctccaagc	acctgacactg	cgagtgcag	gacctcatct	accacatgtc	gcagcccgac	840
gtcagccgc	ggctccacat	cgatgagatc	ctcagccact	cgtggctgca	gcccccccaag	900
cccaaagcca	cgtttctgc	ctccttcaag	agggaggggg	agggcaagta	ccgcgctgag	960
tgcaaactgg	acaccaagac	aggcttgagg	cccgaccacc	ggccgacca	caagcttgg	1020
gccaaaaccc	agcacccggct	gctgggtgg	cccgagaacg	agaacaggat	ggaggacagg	1080
ctggccgaga	cctccaggccc	caaagaccat	cacatctccc	gagctgaggt	ggggaaagca	1140
agcacctagc	atgacaatgg	ccccgttgg	tgtgggtgggg	gtcgggggtt	gggggcatgg	1200
tgcagtcggc	ttcacgtaa	actaagtagg	caggtaggat	ctgaagaagg	cacaggtgca	1260
agtaaaatc	gtcaattaaa	ccactatttt	gattacgttc	cattagcttt	cttccactta	1320
gcagcaaaga	cgttcattac	tgaccaccaa	ataaaccaca	gggtgtgtgc	aagcatcaa	1380
						1381

&lt;210&gt; 19

&lt;211&gt; 3904

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3210906CB1

&lt;400&gt; 19

tattcggggt tcagacccca caatcagaaa tccggaattc ggcagctgtc gccctcgacg 60  
 agggggagga ctggaccgcg aggtcagatt aggttgtcac cccctccct ccaggggagg 120  
 ctteccggc cggccccta ggaaggcga aagccgagga agagggtggca aggggaaagg 180  
 tctccttgc cctctccctg acttggcaga gccgctggag gaccccaggc ggaagcggag 240  
 gcgctgggc accatagtga cccctaccag gcaggcccc actctcaggg ccccccagggg 300  
 ccaccatgcc agctggggc cgggccccga gcctgaagga cccagatgtg gctgagctct 360  
 tcttcaagga tgacccagaa aagctttct ctgacctccg gaaattggc catggcagct 420  
 ttggagccgt atacttgcg cggatgtcc gaaatagtga ggtggtggcc atcaagaaga 480  
 tgccttacag tgggaagcag tccaatgaga aatggcaaga catcatcaag gaggtgcgg 540  
 tcttacagaa gctccggcat cccaaacacca ttctgatccg gggctgttac ctgagggagc 600  
 acacggctt gctggtaatg gatgttgc tggctcaac ttctgacctt cttagaagtgc 660  
 acaagaaacc cctttaggg ctagatgtc cagctgtgac ccacggggcg cttcaggggc 720  
 tggcatatct gcaactccac aacatgtcc ataggatgt gaaggctgga aacatccctc 780  
 tgcagagcc agggtagtg aagcttaggg actttggttc tgcgtccatc atggcacctg 840  
 ccaactcctt cgtgggcacc ccatactgga tggcacccga ggtgatcctg gccatggatg 900  
 aggggcagta cgatggcaaa gtggacgtct ggtccttggg gataacctgc atcgagctgg 960  
 ctgaacggaa accaccgctc tttaacatga atgcgtatc tgccttatac cacattgcac 1020  
 agaacgaatc cccctgtctc cacttcgtct tcagaaaatc cctcaagaca gaccaaccc agagggttctc aattttgtcg 1080  
 gctttgtct cggggagcgg atgcgtgtcg ggagctggac agcgtcagt accgcaatg gaagaagatc ctgttccaag 1140  
 agggggcggcc cggggcttggt tgcacctggc cggggactctg tgcgtccatc atggcaccc 1200  
 tgcgtccctc cagccagagc aagaggagga ggaggaggag aggccctgaa gcccggaga 1260  
 tggccatgtat gcaaggagggg gacacacag tcaacctctc cagctccatt atccaccggc 1320  
 tgccggggtc tgacaaccta tcttgcggcc tgcagccccca acttcggtaa ccgagaccac tttgcacca tccgaaccgc ctcctggc 1380  
 tccaggagca tgacggcggc gacacagcggc cggggacttgc agagtagccca ctcagtgc 1440  
 tcaagccctc cagccagagc aagaggaggg aggagggaga aggccctgaa gcccggaga 1500  
 tggccatgtat gcaaggagggg gacacacag tcaacctctc cagctccatt atccaccggc 1560  
 tgccggggtc tgacaaccta tcttgcggcc tgcagccccca acttcggtaa ccgagaccac tttgcacca tccgaaccgc ctcctggc 1620  
 gctccactt ccaccaccc tttgcacca tccgaaccgc ctcctggc 1680  
 tttgcacca tccgaaccgc ctcctggc 1740  
 tctgcgtgc gggagcgtat gacggctat aagcggatgc 1800  
 ctgtggccc tggagtcaatc gtcgtggatc 1860  
 cggggacttgc aggcgcagcg gctggctt gggcagagg 1920  
 caccaggcca taggtgagaa ggaggcacga gctgcccagg 1980  
 ccgaggagcg gaagttccag cagcacatcc ttggcagca gaagaaggag ctggctgccc 2040  
 tgctggaggc acagaagcg acctacaaatc ttcgcaagga acagctgaag gaggagctcc 2100  
 aggagaaccc cagcactccc agctccagca gtggccaggcg tttgcgtca gtgtcgccag aggacactgtc gcccggcc 2160  
 ttgcgtca gtgtcgccag aggacactgtc gcccggcc 2220  
 tttgcacca tccgaaccgc ctcctggc 2280  
 tttgcacca tccgaaccgc ctcctggc 2340  
 tttgcacca tccgaaccgc ctcctggc 2400  
 tttgcacca tccgaaccgc ctcctggc 2460  
 tttgcacca tccgaaccgc ctcctggc 2520  
 tttgcacca tccgaaccgc ctcctggc 2580  
 tttgcacca tccgaaccgc ctcctggc 2640  
 tttgcacca tccgaaccgc ctcctggc 2700  
 tttgcacca tccgaaccgc ctcctggc 2760  
 tttgcacca tccgaaccgc ctcctggc 2820  
 tttgcacca tccgaaccgc ctcctggc 2880  
 tttgcacca tccgaaccgc ctcctggc 2940  
 tttgcacca tccgaaccgc ctcctggc 3000  
 tttgcacca tccgaaccgc ctcctggc 3060  
 tttgcacca tccgaaccgc ctcctggc 3120  
 tttgcacca tccgaaccgc ctcctggc 3180  
 tttgcacca tccgaaccgc ctcctggc 3240  
 tttgcacca tccgaaccgc ctcctggc 3300  
 tttgcacca tccgaaccgc ctcctggc 3360  
 tttgcacca tccgaaccgc ctcctggc 3420  
 tttgcacca tccgaaccgc ctcctggc 3480  
 tttgcacca tccgaaccgc ctcctggc 3540  
 tttgcacca tccgaaccgc ctcctggc 3600  
 tttgcacca tccgaaccgc ctcctggc 3660  
 tttgcacca tccgaaccgc ctcctggc 3720  
 tttgcacca tccgaaccgc ctcctggc 3780  
 tttgcacca tccgaaccgc ctcctggc 3840

agcaggtctg gggcctgagg cagggagaat gagaggccac ttactggcag gaaggatcg 3900  
gatg 3904

<210> 20  
<211> 1987  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3339024CB1

<400> 20  
gaagaaccc gaggAACAGA cttacccTAG cAACCCCTGGC acctCCAAACC CGACACATGC 60  
tactgctgt gctactgctg ccacCCCTGC tCTGTGGAG agtGGGGGT aAGGAACAGA 120  
aggattacat gctgacaatg tagaaGTCCG tgacGGTGCa ggAGGGCCTG tGTGTCTCTG 180  
tgctttgctc ctttcctac ccccaaaATG gctggactgc ctccgatcca gttcatggct 240  
actgggttccg ggcaggGGAC catgtAAAGCC ggaacattcc agtggCCACa aacaACCCAG 300  
ctcgagcagt gcaggAGGGAG actcgggACc gattccACCT cttGGGGAC ccacAGAAACA 360  
aggattgtac cctgagcATC agagACACCA gagAGAGTGA tgcAGGGACA tacgtcttt 420  
gtgttagAGAG aggAAATATG aaATGGAAATT atAAATATGA ccAGCTCTCT gtGAATGTGA 480  
cagcgtccca ggacctACTG tcaAGATAcA ggCTGGAGGT gCCAGAGTCg gtGACTGTGc 540  
aggagggtct gtgtgtctct gtgcCCTGCA gtgtCCCTTA cccccattac aactggactg 600  
cctctagccc tgTTATGGA tcctGGGTC aggaAGGGGC cgatataccA tgggatattc 660  
cagtggccac aaACACCCCA agtggAAAGAG tcaAGAGGA taccACCGT cgattCCtCC 720  
tcctGGGGGA cccACAGACC aacaACTGt ccctgAGCAT cAGAGATGCC aggaAGGGGG 780  
attcaggGAA gtactACTC caggtggAGA gAGGAAGCAG gAAATGGAAAC tacatATATG 840  
acaagctcC tGTGcatgtg acagCCtGA ctcACATGCC cacTTCTCC atcccGGGGa 900  
ccctggagtc tggccACCCC aggaACCTGA cctgCTCTGT gCCtGGGGC tGTGAACAGG 960  
gjacGCCcC cACGATCACC tggatGGGG cctCCGTGTC ctCCCTGGAC cccACTATCA 1020  
ctcgctccTC gATGCTCAGC ctCATCCAC AGCCCCAGGA cCATGGCACC agcCTCACCT 1080  
gtcaggGTAC cttGcCTGGG gCCGGCGTGA cCATGACCAG ggCTGTCCGA ctCAACATAT 1140  
cctatCCtCC tcGAACATtG accatGACTG tCTTCCAAGG agatGGCACA gcatCCACAA 1200  
ccttGAGGAG cttGCTGGCC tttCAgTCC tggAGGGCca gtcCCTGcAC cttGtCTGT 1260  
ctgtcGACAG caatCCCCCT gCCAGGTGA gCTGGACCTG gGGGAGGCTG accCTGAGCC 1320  
cctCACAGtC ctCGAACCTT gGGGTGCTGG agtGcCTCG agtGcatGTG aaggatGAAAG 1380  
ggaattcAC ctGCGCAGtC cagaACCCtC tagGCTCCCA gCACATTCC ctGAGCCTCT 1440  
ccctgcaAAA cgAGTACACA ggCAAAATGA gGCCTATATC aggAGTGAcG ctAGGGGcAT 1500  
tcggGGGAGC tggAGCCACA gCCCTGtCT tCTGTACTT ctGcatCAtC ttcGTTGTA 1560  
tgaggGCTG cAGGAAGAAA tcggCAAGGC cAGCAGTGGG cgtGGGGAT acAGGcAtGG 1620  
aggacGCAA CGCTGTCTGG ggCTCAGCCT CTCAGGGACC CCTGATTGAA tcccCGGcAG 1680  
atGACAGCC CCCACACCAT gCTCCGCCAG CCCTGGCCAC CCCCTCCCCA gAGGAAGGAG 1740  
agatCCAGtA tgcAtCCtC AGtGCAcACA aAGCAGGGCC tcaGtACCA cAGGAACAGG 1800  
aggCCatCAG cTATGAGtAC tccGAGAtCA ACATCCCCAA GTGAGAAACT GcAGAGACTC 1860  
aggCCtGTT gagGGCtAC gACCCtCCa gCAAAAGAAGC CCAGAGACTGA ttcCtttAGA 1920  
attAAAAGCC ctCCatGtGtG tgcaACGGG gatCCACTAG ttaAGAGCGG CGCACCCCG 1980  
tgccccCT 1987

<210> 21  
<211> 3925  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 4436929CB1

<400> 21  
ccgtcctcga ggcgAGGGAGA gtaccGGGCC ggCCCGGCTG ccgcgcgAGG AGCGCGGTG 60  
gcggcctggT ctgcggCTGA gataCACAGA gCGACAGAGA cATTATTGT tATTTGTTTT 120  
ttggTGGCAA aaAGGGAAAa tggcGAACGA ctCCCTGCA AAAAGTCTGG tggACATCGA 180  
cctCCtCCtCC ctgcGGGATC ctGCTGGGT tttGAGCTG gtGGAAGTGG ttggAAATGG 240  
cacCTatGGA caAGTCTATA agggTCGACA tGTTAAACAG ggtcAGTtGG cAGCCATCAA 300  
agttatGGAT gTCactGAGG atGAAGAGGA AGAAATCAAAGA CTGGAGATAA atATGCTAAA 360  
gaaataCTtC catCACAGAA ACATTGCAAC ATATTATGGT gCTTtCATCA AAAAGAGCCC 420  
tccAGGACAT gatGACCAAC tctGGCTGT tatGGAGtC tGtGGGGCTG ggtCCATTAC 480  
agacCTTGTG aAGAACACCA aAGGGAACAC ACTCAAAGAA gACTGGATCG cttACATCTC 540

cagagaaaatc	ctgaggggac	tggcacatct	tcacattcat	catgtgattc	acggggatat	600
caagggccag	aatgtgttgc	tgactgagaa	tgcagaggtg	aaacttggtg	actttgggt	660
gagtgtctag	ctggacagga	ctgtgggchg	gagaataacg	ttcataggca	ctccctactg	720
gatggctct	gaggtcatcg	cctgtatga	gaacccagat	gccacctatg	attacagaag	780
tgatcttgg	tctgtggca	ttacagccat	tgagatggca	gaagggtgctc	cccccctctg	840
tgacatgtcat	ccaatgagag	cactgtttct	cattccaga	aaccctctc	cccggtctgaa	900
gtcaaaaaaa	tggcgaaga	agtttttag	ttttagaa	gggtgcctgg	tgaagaatta	960
catgcagcgg	ccctctacag	agcagctttt	gaaacatct	tttataaggg	atcagccaaa	1020
tgaaggcga	gttagaatcc	agcttaaggaa	tcataatagat	cgtaccaggaa	agaagagagg	1080
cgagaaagat	gaaaactgagt	atgagtagac	tgggagtgag	gaagaagaggg	aggaagtgc	1140
tgaacaggaa	ggagagccaa	gttcattgt	gaacgtgcct	ggtgagtctc	ctcttcgccc	1200
agatttctg	agactgcagc	aggagaacaa	ggaacgttcc	gaggcttcc	ggagacaaca	1260
gttactacag	gagcaacagc	tccgggagca	ggaagaatat	aaaaggcaac	tgctggcaga	1320
gagacagaag	cggattgagc	agcagaaaaga	acagaggcga	cggctagaag	agcaacaag	1380
gagagagcgg	gaagctagaa	ggcagcagga	acgtgaacag	cgaaggagag	aacaagaaga	1440
aaagaggcgt	ctagaggagt	tggagagaag	gcbcggaaaga	gaagaggaga	ggagacgggc	1500
agaagaagaa	aagaggagag	ttgaaagaga	acaggagat	atcaggcgac	actagaaga	1560
ggagcagcgg	catttggaa	tcctcagaa	cgcagctgtc	caggagcagg	ccatgttact	1620
gcatgaccat	aggaggccgc	acccgcagca	ctcgcagcag	ccgcccaccac	cgcagcagg	1680
aaggagcaag	ccaaagttcc	atgctcccg	gcccggaaagcc	cactacgagc	ctgctgaccg	1740
agcgcgagag	gttggaaagata	gatttaggaa	aactaaccac	agctcccctg	aagccagtc	1800
taagcagaca	ggcagagttat	tggagccacc	agtgcctcc	cgatcagagt	cttttccaa	1860
tggcaactcc	gagtctgtgc	atccgcctt	gcagagacca	gcggagccac	aggttctgt	1920
gagaacaaca	tctcgctccc	ctgttctgtc	ccgtcgagat	tcccccactgc	agggcagtgg	1980
gcagcagaat	agccaggcag	gacagagaaa	ctccaccaggc	agtattgagc	ccaggcttct	2040
gtggggagaga	gttggagaagc	tggtcccgg	acctggcagt	ggcagctct	cagggtccag	2100
caactcaggaa	tcccgcccc	ggtctcaccc	tgggtctcag	agttggctccg	ggggacgctt	2160
cagagtgaga	tcatcatcc	agtcgtacgg	ctctccatct	cagcgcctgg	aaaatgcagt	2220
aaaaaaaccc	gaagataaaaa	aggaagttt	cagacccttc	aaggctgctg	gcgaagtgg	2280
tctgaccgca	ctggccaaag	agcttcgagc	agtggaaagat	gtacggccac	ctcacaagg	2340
aacggactac	tcctcatcca	gtgaggagtc	ggggacgacg	gatgaggagg	acgacatgt	2400
ggagcaggaa	ggggctgacg	agtccaccc	aggaccagag	gacaccagag	cagcgtcatc	2460
tctgaatttgc	agcaatgggt	aaacggaaatc	tgtaaaaacc	atgattgtcc	atgatgtatgt	2520
agaaaatgtag	cggggccatca	ccccatccaa	ggagggcaact	ctaatcgtc	gccagactca	2580
gtccgtctgt	agcacactcc	agaaacacaa	atcttcctcc	tccttccatc	cttttataga	2640
coccagattt	ctacagattt	ctccatctag	cggaaacaaaca	gtgacatctg	tgttgggatt	2700
ttcctgtgt	gggatgagac	cagaagccat	aaggcaagat	cctacccggaa	aggctcagt	2760
ggtcaatgtg	aatcctacca	acactaggcc	acagagtgcac	acccgggaga	tcgtaaata	2820
caagaagagg	ttaactctg	agattctgt	tgctgcctt	tggggagtg	atttgctagt	2880
gggtacagag	agtggcctga	tgctgttgg	cagaagtggc	caagggaaagg	tctatccct	2940
tatcaaccga	agacgattt	aacaaatgg	cgtacttgag	ggcttgaatg	tcttggtgac	3000
aatatctggc	aaaaaggata	agttacgtgt	ctactat	tcctgggtta	gaaataaaat	3060
acttcacat	gatccagaag	ttgagaagaa	gcagggtatgg	acaaccctgat	gggatttgg	3120
aggatgtgt	cattataaa	ttgtttaaaa	tgaaagaatc	aaatttctgg	tgattgtctt	3180
gaagaggctt	gttggaaagtct	atgcgtgggc	accaaaaggca	tatcacaat	ttatggcctt	3240
taagtcat	ggagaatttgg	tacataagcc	attactgtgt	gatctactg	ttgaggaagg	3300
ccagaggttgc	aaagtgtatct	atggatcttgc	tgctggattt	catgtgtt	atgtggattc	3360
aggatcagtc	tatgacattt	atctaccaac	acatatccag	tgttagcatca	aacccatgc	3420
aatcatcatc	ctccccaaata	cagatggaa	ggagcttctg	gtgtgtat	aagatgaggg	3480
gtttagtgc	aacacatatg	gaaggtatcac	caaggatgt	gttctacagt	ggggagagat	3540
gcctacatca	gtagcatata	ttcgatccaa	tcagacaatg	ggctggggag	agaaggccat	3600
agagatccga	tctgtggaaa	ctggtcaat	ggatgggtgt	ttcatgcaca	aaagggtctca	3660
aagactaaaa	tctttgtgt	aacgcaatga	caaggtgttc	tttgcctctg	ttcggcttgg	3720
tggcagcagt	cagggtttat	tcatgaccc	aggcaggact	tctttctgt	gtgttagaa	3780
gcagtgtgt	ccagggattt	ctggcctcca	gagtcttcaa	gatcctgaga	acttgaatt	3840
ccttgcata	ggagctcgga	gctgcaccga	gggcaaccag	gacagctgt	tgtcagacc	3900
tcatgtgttgc	ggttctctcc	cctcc				3925

<210> 22  
<211> 1210  
<212> DNA  
<213> *Homo sapien*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 5046791CB1

<400> 22

ttacagggtca	tctaccccta	taccccacaa	aatgacgatg	agctggagct	ggtccccggg	60
gacttcatct	tcatgtctcc	aatggagcag	accagcacca	gcgagggttg	gatctatggc	120
acgtccttaa	ccacccggctg	ctctggactc	ctcctgagaa	ttacattacc	aaggctgatg	180
aatgcagcac	ctggatattt	catggtttctt	attcaatctt	aaatacatcg	tcatccaact	240
ctctcacgtt	tggggatgga	gtattggaga	ggcggcccta	tgaggaccag	gggctcgaaa	300
agacgactcc	tcttactatc	atctgcagc	ccatgcagcc	gctgagggtc	aacagccagc	360
ccggccccca	gaagcgatgc	ctttttgtgt	gtcggcatgg	tgagaggatg	gatgttgggt	420
ttgggaagta	ctggctgtcc	cagtgcctcg	atgccaaagg	ccgctacata	cgcaccaacc	480
tgaacatgcc	tcatagttt	cctcagcgg	gtggtggtt	ccgagattac	gagaaagatg	540
ctcccatcac	tgtgtttgga	tgcatacaag	caagactagt	gggtgaagcc	ttatttagaga	600
gcaataccat	tatcgatcat	gtctattgtct	ccccgtccct	tcgctgcgtt	cagactgcac	660
acaatataat	gaaagggtta	caacaagaaa	atcaattgaa	gatcgtgt	gagcccggt	720
tatttgatgt	gacaaaatgg	gttgctggg	gcacattacc	tgcataggata	cctccatcag	780
agtttagtgc	agccaaatctg	agtgttata	caacccatcag	acccacatt	ccaatcagca	840
atttagttgt	ttcagaatcc	tatgataactt	atatacgtag	aagtttccaa	gtaacaaaag	900
aaataataag	tgaatgtaaa	agtaaaggaa	ataacatctt	gattgtggcc	cacgcatttt	960
cccttgaagc	gtgtacctgc	caacttcagg	gctgtcacc	tcagaactcc	aaggactcg	1020
tacaaatgtt	ccgaaaagatc	ccatatctgg	gattttgtt	ctgtgaagaa	ttaggagaaa	1080
ctggaaatatg	gcagctgaca	gatccaccaa	tccttcctct	taccatgga	ccaactgggg	1140
gcttcaactg	gagagagacc	ttgcttcaag	aataaaccat	accagtgaac	aagaaggaaa	1200
aaaaaaaaaaa						1210

<210> 23

<211> 1521

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1416174CB1

<400> 23

ggcacgggtgc	tgggcctcgc	cacactgcac	acctaccgcg	gacgcctctc	ctacccccc	60
gccactgtgg	aacccgcctc	gcccacccct	gccccatagcc	tgcctcgtgc	caagtcggag	120
ctgaccctaa	ccccagaccc	agccccccccc	atggcccact	caccctgc	tcgttctgt	180
tctgacccctc	ctcttccct	gccccagcct	gcctggcc	ctcctggctc	gccagaaccc	240
ctgcccattcc	tgtccctcaa	cggtggggc	ccagagctgg	ctggggactg	gggtggggct	300
ggggatgttc	cgctgtcccc	ggacccactg	ctgtcttcc	ctcctggctc	tcccaaggca	360
gctctacact	cacccgtc	cgaaggcccc	ccgttaattcc	cccatctct	gggctccac	420
ttccccaccc	tgtatggccgg	gtagggcc	ccacctgcgg	ccggccgc	cacctgtc	480
ctccgctgg	caccccgctg	cccccaact	gggtgacgct	ggagggggac	tttgtgtca	540
tgttggccat	ctcgccccag	cacccatggc	ctgacccctg	ggcaagctccg	catgcgcgt	600
tcgacgacgg	cctgggtgcac	ctgtgttgg	tgcgtacgg	catctcgcgg	gctgcgtc	660
tgccgccttt	cttggccatg	gagcgttgc	gccacttc	cctgggtgt	ccgcagctgg	720
gctacggccg	ggcccggtgc	ttccgcctag	agccgctc	accacgcggc	gtgctcacag	780
tggacgggg	gcaggtggag	tatggccgc	tacaggcaca	gatgcaccct	ggcatcggt	840
cactgctc	tgggcctct	ggctgcccgg	ggcggggagcc	ctgaaactaa	acaagcttgg	900
tacccgcgg	gggcggggcc	tacattccaa	tggggcgag	ctgagctagg	gggtgtggcc	960
tggctgtctag	agttgtgggt	gcagggggcc	tggcccccgtc	tcagattgc	gctcgcttcc	1020
atgggaccag	acgtgtatgt	ggaagggtgg	cgtcgatc	gtttaagaga	aatggggtcg	1080
tcccgagggt	agtgcctgtat	caatggggc	ggggcctggc	gtcgatctg	ggggccccc	1140
tacggggcag	ggctcagtc	tgacgttgc	cacccgttcc	taccggcca	ggatggctga	1200
gggcggagtc	tattttacgc	gtcgccca	gacaggaccc	ggaatgtact	ggctggggta	1260
ggcctcagt	agtcggccgg	tcaggcccc	cagccctcgcc	ccatccactc	cggtgcctcc	1320
atttagctgg	ccaatcagcc	caggagggc	aggttcccg	gggcggcgc	taggatttgc	1380
actaatgttc	ctctccccgc	gggtggggc	ggggaaattc	atatccctg	ttcgctctcat	1440
ggcgctcc	cgtccccaa	ctaaaaagca	attgaaaagg	tctatgcaat	aaaggcagtc	1500
gcttcattcc	tctaaaaaaa	a				1521

<210> 24

<211> 1640

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 3244919CB1

<400> 24

gcagcgccgc	ggcgccccgg	ggctcgccgc	ccccggccgc	cgcgcccc	gccggctccg	60
acgcgccttc	ggccctgccc	ccgcccgcgt	ctggccagcc	ccggggcccg	gactcggcg	120
atgtccgcgc	cgagccgcgc	ccctgtttc	agtggagcaa	gtggagaag	aggatgggt	180
cgtccatgtc	ggcgccacc	gcgcggaggc	cgtgtttga	cgacaaggag	gacgtgaact	240
tcgaccactt	ccagatcctt	cgggcatttgc	ggaaggcgag	ctttggcaag	gtgtgcatttgc	300
tcgagaagcg	ggacacggag	aagatgtacg	ccatgaagta	catgaacaag	cagcagtgc	360
tcgagcgcga	cgaggtccgc	aacgtttcc	gggagcttgg	gatcctgcag	gagatcgagc	420
acgtcttcct	ggtaaacctc	tggtacttct	tccaggacga	ggaggacatg	ttcatggtcg	480
tggacctgt	actggggcggg	gacctgcgt	accacctgca	gcagaacgtg	cagttctccg	540
aggacacggt	gaggctgtac	atctgcgaga	tggcacttggc	tctggactac	ctgcgcggcc	600
agcacatcat	ccacagagat	gtcaaggctg	acaaacatttct	cctggatgag	agaggacatg	660
cacacctgac	cgacttcaac	attggccacca	tcatcaagga	cggggagcgg	gcgacggcat	720
tagcggcac	caagccgtac	atggctccgg	agatcttcca	ctctttgtc	aacggcggga	780
ccggctactc	cttcgaggtg	gactgggt	cgggtgggggt	gatggcctat	gagctgctgc	840
gaggatggag	gccctatgac	atccacttca	gcaacgcgcgt	ggagtccttgc	gtcagctgt	900
tcagcacccgt	gagcgtccag	tatgtcccc	cgtggtccaa	ggagatggtg	gccttgcgtc	960
ggaagctct	cactgtgaac	cccgagcacc	ggctctccag	ccttcaggac	gtgcagggcag	1020
ccccggcgct	ggccggcggt	ctgtgggacc	acctgagcga	gaagagggtg	gagccggct	1080
tcgtccccaa	caaaggccgt	ctgcactgcg	accccacctt	tgagctggag	gagatgatcc	1140
tggagtcac	gccctctgcac	aagaagaaga	agcgccctggc	caagaacaag	tcccgggaca	1200
acagcaggga	cagctcccg	tccggcccac	ggagcaagtc	caagccatcc	acccagaggc	1260
aaggggatgt	ggcccttggca	tcctcggtt	tgggagaatg	actatcttca	agactgcctc	1320
gatgccatcc	agcaagactt	cgtgatTTT	aacagagaaa	agctgaagag	gagccaggac	1380
ctcccgagggg	agccctctccc	cgccccctgag	tccaggatg	ctgcggagcc	tgtggaggac	1440
gaggcggaaac	gctccgcctt	gcccattgtgc	ggccccattt	gcccctcgcc	cgggagcggc	1500
taggccccgg	cgccctgtgt	cctcaccctt	tgagctgtt	tggagactcg	gctgccagag	1560
ggagggccat	ggggcgaggc	ctggcattca	cgttcccacc	cagcctggct	ggcggtgccc	1620
acagtgccttc	ggacacattt					1640